Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

PÓLYA TYPE DISTRIBUTIONS IV. SOME PRINCIPLES OF SELECTING A SINGLE PROCEDURE FROM A COMPLETE CLASS[1]

By Samuel Karlin

Stanford University

0. Introduction. In previous publications [1], [2], and [3], various aspects of decision theory in which the underlying distributions are Pólya type have been studied. For example, complete classes of decision procedures were determined, all Bayes procedures were characterized, and the problem of admissibility was investigated as related to various kinds of loss functions.

Usually the minimal complete class of decision procedures, to which the statistician would obviously restrict himself in practical application, is still quite large. Consequently, without any additional knowledge or further conditions, it is a hopeless task to justify preferring any given admissible procedure to another. It is therefore of importance to introduce new criteria which will single out a procedure for use. It is the object of this paper to discuss some further principles which select a single statistical procedure from the class of all "monotone" procedures.

In the $n = 2$ action problem (essentially the testing problem) some of the classical principles used to determine a single admissible procedure for use are related to the concepts of unbiasedness, maximum likelihood, invariance, minimax, etc. These principles have received much attention and their justification and relevance are well understood for the parametric testing problem. For a detailed analysis of these classical concepts in the case of two action problems when the underlying distributions are Pólya type, the reader is referred to [1]. Our present discussion deals with the extension and analysis of some of these principles to the n-action problem. In the sense that the estimation problem may be obtained as a limit of finite action problems, the ideas here shed further light on the estimation problem.

The language and notation we use is that of the introduction of the previous paper [3]. However, a knowledge of the results of [3] is not necessary for an understanding of the present discussion although a reading of the introduction would more than provide sufficient familiarity with the terminology to be used here as well as a general background for Pólya type distributions. Henceforth, we assume that the notation of this manuscript is that of [3]. Nevertheless, for clarity of exposition, we review briefly some of the main quantities to be used.

Let the distribution of the observed real random variable X (usually a sufficient

---

Received March 25, 1957.

[1] Research sponsored by the Office of the Naval Research Nonr 225 (21) (NR 042-993). Research on the paper "Pólya-Type Distributions III: Admissibility of Multi-Action Problems," in the December, 1957 issue of these *Annals*, was done under the same contract.

statistic), depending on the unknown parameter ω ($\omega \varepsilon \Omega$, an interval of the real line), have the form

$$P(x, \omega) = \int_{-\infty}^{x} p(\xi, \omega)\, d\mu(\xi), \tag{1}$$

where the density $p(\xi, \omega)$ possesses a monotone likelihood ratio (Pólya type 2) and μ is a countably additive measure defined at least for the Borel field of sets containing the open subsets of the real line. Occasionally, we shall assume the stronger condition that the density is Pólya type 3.

The main transformation property of Pólya type 2 densities used in our analysis is as follows: If $g(x)$ changes sign at most once (say from negative to positive values), then

$$h(\omega) = \int g(x)p(x, \omega)\, d\mu(x)$$

changes sign at most once. Moreover, if $h(\omega)$ does indeed change signs, then it must change in the same direction as g, i.e., from negative to positive. For a thorough discussion of these properties the reader is referred to [2].

There are n possible actions, and $L_i(\omega)$ ($i = 1, \cdots, n$) represents the measure of the loss when taking action i and ω is the state of nature. We require that the set

$$S_i = \{\omega \mid L_i(\omega) < L_j(\omega), j \neq i\} = (\omega_{i-1}^0, \omega_i^0)$$

where the ω_i^0 satisfy

$$-\infty = \omega_0^0 < \omega_1^0 < \omega_2^0 < \cdots < \omega_n^0 = \infty.$$

The set S_i represents the set of ω values where action i is favored if the state of nature were known. Also, we assume that $L_i(\omega) - L_{i+1}(\omega)$ has exactly one sign change which must occur at ω_i^0.

We shall assume throughout what follows that the loss functions $L_i(\omega)$ and the density $p(x, \omega)$ satisfy sufficient smoothness conditions to guarantee the existence of all integrals involving these quantities and to justify all differentiation operations. In most particular examples these smoothness requirements can be readily verified.

A statistical procedure is an n-tuple

$$\varphi(x) = (\varphi_1(x), \cdots, \varphi_n(x)),$$

where $\varphi_i(x)$ is interpreted as the probability of taking action i when observing x. A "monotone" procedure is characterized by a tuple

$$(x_1, x_2, \cdots, x_{n-1}; \lambda_1, \cdots, \lambda_{n-1})$$

where $x_1 \leqq x_2 \leqq \cdots \leqq x_{n-1}$, $0 \leqq \lambda_i \leqq 1$. Explicitly, when the x_i are distinct, then

$$\varphi_i(x) = \begin{cases} 1 & \text{if } x_{i-1} < x < x_i, \\ 0 & \text{if } x < x_{i-1}, x > x_i, \\ \lambda_i & \text{if } x = x_i, \\ 1 - \lambda_{i-1} & \text{if } x = x_{i-1}, \end{cases} \qquad i = 1, \cdots, n,$$

and by definition $x_0 = -\infty$, $\lambda_0 = 0$, $x_n = +\infty$, $\lambda_n = 1$. In the case where some of the x_i coincide then appropriate changes in the form of the definition of $\varphi_i(x)$ at the values x_i must be made. If the measure μ of (1) has no atoms (jumps), then a monotone procedure is fully specified (up to equivalence almost everywhere with respect to μ) by the critical values $(x_1, x_2, \cdots, x_{n-1})$. For the sake of simplicity of exposition, we restrict ourselves henceforth to the case of a continuous distribution. However, we remark in passing that all of the results of this paper may be extended, subject to suitable modifications, to the general case where we allow the measure μ to possess atoms. The risk corresponding to any given strategy $\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_n)$ is given by the expression

$$\rho(\omega, \varphi) = \int p(x, \omega) \left\{ \sum_{i=1}^{n} L_i(\omega)\varphi_i(x) \right\} d\mu(x). \tag{2}$$

The collection of all monotone procedures constitutes a complete class [4]. When the loss functions satisfy additional assumptions, then all non-degenerate monotone procedures are also admissible [3].

The set of all monotone strategies $\mathfrak{M}$ form an $n - 1$ dimensional family in the sense that they depend on the $n - 1$ critical values which determine the procedures. Our problem, in choosing a specific strategy from $\mathfrak{M}$, is in essence finding $n - 1$ conditions which will cut the class $\mathfrak{M}$ down to a unique member. Alternatively, we could impose some global restrictions which also single out a monotone procedure. For instance, if an a priori distribution of nature $F(\omega)$ is known to be meaningful, then the Bayes procedure with respect to F determines a specific monotone procedure. [See [3], [5].] The assumption of the existence of F is often hard to justify and appears contrived.

Another global condition frequently followed is to choose a monotone minimax procedure. However, minimax procedures are often very unreasonable on the basis of statistical intuition and there exists feeling that minimax philosophy is in general too conservative and unrealistic. Of course, modifications of the minimax principle lead to the so-called regret principles. Various complications appear also for the case of the criteria of minimax regret [6].

A third method for choosing a monotone procedure is inherent in the construction of complete classes as introduced in [4]. Suppose that for a given problem there has been in use a common or accepted mode of action which is not a monotone procedure. Then, there exists at least one monotone procedure which improves everywhere on it for the decision problem of more than two actions. If the original procedure is described by an n-tuple of functions $\varphi = (\varphi_1, \cdots, \varphi_n)$, then any monotone procedure $\varphi^0 = (\varphi_1^0, \cdots, \varphi_n^0)$ (and there is at least one) which satisfies

$$\int_{-\infty}^{\infty} p(x \mid \omega) \left[\sum_{j=1}^{i} \varphi_j^0(x) - \sum_{j=1}^{i} \varphi_j(x) \right] d\mu(x) \begin{cases} \geqq 0 & \text{for} \quad \omega \leqq \omega_i^0 \\ \leqq 0 & \text{for} \quad \omega \geqq \omega_i^0 \end{cases}$$

improves on φ.

This method is constructive. That is, for any non-monotone procedure in use we can explicitly exhibit a monotone procedure which yields a smaller risk uniformly for any choice of the state of nature ω. The apparent disadvantage to this idea is that it involves only an improvement relative to a given non-monotone procedure and sheds no light on the intrinsic question of selecting a specific monotone procedure from the class $\mathfrak{M}$.

In this study we will analyze three principles of selecting a single monotone procedure from $\mathfrak{M}$. The first represents an extension of the maximum likelihood estimate to the circumstance of the n-action problem. The monotone test obtained in this case has a lot of intuitive appeal and will be referred to as the maximum likelihood procedure.

The following section examines another approach called the principle of maximum probabilities (abbreviated M.P.). This principle, as well as the maximum likelihood procedure, does not depend on the specific values of the loss functions but rather on the preference regions $S_i = (\omega_{i-1}^0\ \omega_i^0)$. Any other loss function satisfying the properties of a monotone preference pattern and giving rise to the same preference sets S_i will possess the same class of monotone procedures obeying the principle of M.P.

The precise description of this principle is as follows: A decision procedure which is defined by an n-tuple of functions $\varphi = (\varphi_1, \cdots, \varphi_n)$ is said to have the property of maximum probabilities (φ has M.P.) if for every i

$$h_i(\omega') \geqq h_i(\omega'') \qquad \text{for any } \omega' \text{ in } S_i, \omega'' \not\in S_i, \tag{3}$$

where

$$h_i(\omega) = \int \varphi_i(x)p(x, \omega)\, d\mu(x). \tag{4}$$

For the case of two actions a procedure φ has the property of M.P. if and only if φ is unbiased in the classical sense. Therefore, this principle may be considered to be a generalization to the case of n actions of the concept of unbiasedness. The quantity $h_i(\omega)$ may be interpreted as the unconditional probability for the procedure φ of taking action i when the state of nature is ω. The condition (3) states that $h_i(\omega)$ is larger when ω is in S_i than when ω is outside S_i. This last property is the reason for the name, principle of maximum probabilities.

It will be shown that there always exist monotone procedures having the property of M.P. for the case of $n \leqq 5$ actions. In fact, we shall exhibit a one parameter family of such procedures. When $n > 5$, in general there ceases to exist such monotone procedures.

The final principle investigated is the principle of unbiasedness (in the sense of Lehmann [7]). A decision procedure φ is said to be risk unbiased with respect to the loss functions L_i if $E_\theta[L(\omega, \varphi(x))] \geqq E_\theta[L(\theta, \varphi(x))]$ for all ω and θ, where $E_\theta(\cdot)$ denotes the expected value given that the state of nature is θ, and

$$L(\omega, \varphi(x)) = \sum L_i(\omega)\varphi_i(x).$$

For the case of two actions, this definition reduces to the usual concept of unbiasedness. This principle of unbiasedness differs from the principle of M.P. in that the former depends in a very crucial way on the magnitudes of the loss functions while the latter depends only on the preference regions. We shall prove that if $L_j(\omega) = L_{ij}$ for ω in S_i and the L_{ij} satisfy suitable assumptions, then there exists a unique admissible monotone procedure unbiased in the sense of Lehmann. The method of proof of the existence will in effect be constructive. In general, risk unbiased procedures need not exist.

ACKNOWLEDGEMENT. I wish to thank Mr. R. Miller for his help in the preparation of this manuscript.

1. Maximum likelihood principle. We assume throughout this section that the density $p(x, \omega)$ of (1) has a strict monotone likelihood ratio and further that $p(x, \omega)$ possesses continuous second order partial derivatives. The fact that p is of Pólya type 2 implies (see [2]) that

$$\begin{vmatrix} p(x, \omega) & \dfrac{\partial}{\partial \omega} p(x, \omega) \\ \dfrac{\partial}{\partial x} p(x, \omega) & \dfrac{\partial^2}{\partial x \partial \omega} p(x, \omega) \end{vmatrix} \geqq 0 \tag{5}$$

for all x and ω. An additional assumption is imposed to the effect that the inequality of (5) is strict for all x and ω. Finally, we assume that for each x in X the equation

$$\frac{\partial}{\partial \omega} p(x, \omega) = 0 \tag{6}$$

has a unique solution, $\omega = \omega(x)$, which is a differentiable function of x. These assumptions are not as stringent as may appear offhand. A wide class of distributions, including the exponential family ($p(x, \omega) = e^{\omega x}\beta(\omega)$), the noncentral t, the noncentral χ^2, etc., fulfills these requirements. For the exponential family, $\omega(x)$ is the solution of the equation $-\beta'(\omega)/\beta(\omega) = x$.

LEMMA 1. *$\omega(x)$ is a strictly increasing function of x.*

Proof. Differentiating Eq. (6) with respect to x leads to

$$\frac{\partial^2 p(x, \omega(x))}{\partial x \partial \omega} + \frac{\partial^2 p(x, \omega(x))}{\partial \omega^2}\, \omega'(x) = 0. \tag{7}$$

By assumption,

$$\begin{vmatrix} p(x, \omega(x)) & \dfrac{\partial}{\partial \omega} p(x, \omega(x)) \\ \dfrac{\partial}{\partial x} p(x, \omega(x)) & \dfrac{\partial^2}{\partial x \partial \omega} p(x, \omega(x)) \end{vmatrix} > 0,$$

which implies $\partial^2 p(x, \omega(x))/\partial x \partial \omega > 0$ because of (6). Since $p(x, \omega)$ assumes a maximum at $\omega = \omega(x)$, $\partial^2 p(x, \omega(x))/\partial \omega^2 \leqq 0$. Thus from (7), $\omega'(x) > 0$.

As x varies over the sample space X, $\omega(x)$ varies over the whole Ω interval. Suppose not; then there exists an ω_0 such that ω_0 is not the upper endpoint of

Ω and for $\omega > \omega_0$, $\partial p(x, \omega)/\partial\omega < 0$ for all x. (Or similarly for the lower end of Ω.) But this contradicts the fact that for all $\omega \varepsilon \Omega$,

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\omega} p(x, \omega)\, d\mu(x) = \frac{\partial}{\partial\omega} \int_{-\infty}^{\infty} p(x, \omega)\, d\mu(x) = 0.$$

Since $\omega(x)$ is a $1 - 1$ strictly monotonic mapping of X onto Ω, the inverse function ω^{-1} is well-defined. Set $x_i^0 = \omega^{-1}(\omega_i^0)$, $i = 1, \cdots, n - 1$. The maximum likelihood principle dictates that the monotone procedure which should be used is the one defined by the critical numbers $(x_1^0, \cdots, x_{n-1}^0)$. For $x \varepsilon (x_{i-1}^0, x_i^0)$, take action i, $i = 1, \cdots, n$, $x_0^0 = -\infty$ and $x_n^0 = +\infty$. This principle has the feature that for any observed x the proper action i is taken whose corresponding interval $(\omega_{i-1}^0, \omega_i^0)$ includes the maximum likelihood estimate of ω. In less precise language, that action is taken which is most likely.

2. Principle of maximum probabilities (M.P.). The principle of maximum probabilities is one type of extension of the concept of unbiasedness in hypothesis testing. Consider the n-action problem defined by the points $-\infty = \omega_0^0 < \omega_1^0 < \cdots < \omega_n^0 = +\infty$ in which action i is preferred in the interval $S_i = (\omega_{i-1}^0, \omega_i^0)$. A decision procedure which is defined by an n-tuple of functions $\varphi = (\varphi_1, \cdots, \varphi_n)$ is said to have the property M.P. if for every i, $h_i(\omega') \geqq h_i(\omega'')$ for any $\omega' \varepsilon S_i$, $\omega'' \not\varepsilon S_i$, where

$$h_i(\omega) = \int_{-\infty}^{\infty} \varphi_i(x) p(x, \omega)\, d\mu(x).$$

Our object is to try to establish the existence of monotone procedures possessing the property of M.P.

It is necessary in studying this concept to assume that the density $p(x, \omega)$ is strictly Pólya type 3, and that the equation $\partial p(x, \omega)/\partial\omega = 0$ is well-defined and has a unique solution $\omega = \omega(x)$ for each value of x. For any constants $a < b$ it is tacitly assumed that differentiation with respect to ω is valid inside the integral sign of

$$\int_a^b p(x, \omega)\, d\mu(x).$$

Also, assume that μ is a continuous measure without discrete mass points whose spectrum is an interval. This last assumption is not essential but without it additional care must be taken in handling randomizations and the lack of uniqueness of various quantities caused by gaps in the spectrum.

For the purpose of exposition our analysis is divided into a series of lemmas.

A randomized strategy is now defined by $n - 1$ points $(x_1, \cdots, x_{n-1})$. Let $i (i = 1, \cdots, n - 2)$ be fixed for the moment and define $(x_i(\alpha), x_{i+1}(\alpha))$ by the equations

$$\begin{aligned} h_{i+1}(\omega_i^0) &= \int_{x_i}^{x_{i+1}} p(x, \omega_i^0)\, d\mu(x) = \alpha, \\ h_{i+1}(\omega_{i+1}^0) &= \int_{x_i}^{x_{i+1}} p(x, \omega_{i+1}^0)\, d\mu(x) = \alpha. \end{aligned} \tag{8}$$

$(x_i(\alpha), x_{i+1}(\alpha))$ are uniquely defined since by Theorem 3 of [1] there is a unique monotone strategy which improves on the non-monotone strategy $\varphi(x) \equiv \alpha$. Moreover, it is clear that $h_{i+1}(\omega') \geqq h_{i+1}(\omega'')$ for any $\omega' \varepsilon S_{i+1}$ and $\omega'' \not\varepsilon S_{i+1}$ when (8) is satisfied.

LEMMA 2. *$x_i(\alpha)$ is a monotone decreasing and $x_{i+1}(\alpha)$ is a monotone increasing function of α.*

Proof. From (8),

$$\int_{x_i(\alpha)}^{x_{i+1}(\alpha)} [p(x, \omega_i^0) - p(x, \omega_{i+1}^0)]\, d\mu(x) = 0 \tag{9}$$

for all α. Since $p(x, \omega)$ is strictly Pólya type 3, $p(x, \omega_i^0) - p(x, \omega_{i+1}^0)$ has at most one zero; by (9) it has at least one. In order that the relation (9) be preserved for all α, either $x_i(\alpha)$ increases and $x_{i+1}(\alpha)$ decreases, or $x_i(\alpha)$ decreases and $x_{i+1}(\alpha)$ increases, as α increases. It is clear from (8) that the latter must hold.

It also follows from the variation diminishing properties of the density $p(x, \omega)$ [2] that

$$h_1(\omega) = \int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x)$$

is a monotone decreasing function of ω, and

$$h_n(\omega) = \int_{x_{n-1}}^{\infty} p(x, \omega)\, d\mu(x)$$

is a monotone increasing function of ω for any x_1 and x_{n-1} respectively.

Consider $x_i(\alpha)$ and $x_{i+1}(\alpha)$, which are defined by the equations $h_{i+1}(\omega_i^0) = \alpha = h_{i+1}(\omega_{i+1}^0)$, and $x'_{i+1}(\alpha)$ and $x'_{i+2}(\alpha)$, which are defined by $h_{i+2}(\omega_{i+1}^0) = \alpha = h_{i+2}(\omega_{i+2}^0)$. Then

LEMMA 3. *For all α, $x_i(\alpha) < x'_{i+1}(\alpha)$ and $x_{i+1}(\alpha) < x'_{i+2}(\alpha)$, $i = 1, \cdots, n-3$.*

Proof. Let

$$I_{[a,b]} = \begin{cases} 1, & a \leqq x \leqq b, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose $x_i(\alpha) \geqq x'_{i+1}(\alpha)$. Then $I_{[x_i, x_{i+1}]} - I_{[x'_{i+1}, x'_{i+2}]}$ is always of one sign or at worse changes sign from $-$ to $+$. But

$$\int_{-\infty}^{\infty} [I_{[x_i, x_{i+1}]} - I_{[x'_{i+1}, x'_{i+2}]}] p(x, \omega)\, d\mu(x) \begin{cases} > 0 & \text{for} \quad \omega < \omega_{i+1}^0 \\ = 0 & \text{for} \quad \omega = \omega_{i+1}^0 \\ < 0 & \text{for} \quad \omega > \omega_{i+1}^0 \end{cases} \tag{10}$$

which is an impossibility in that it changes sign in the wrong direction [2] so $x_i(\alpha) < x'_{i+1}(\alpha)$.

Suppose $x_{i+1}(\alpha) \geqq x'_{i+2}(\alpha)$. Then $I_{[x_i, x_{i+1}]} - I_{[x'_{i+1}, x'_{i+2}]}$ is always of one sign which contradicts (10).

As $\alpha \to 1$, $x_i(\alpha), x'_{i+1}(\alpha) \to -\infty$ and $x_{i+1}(\alpha), x'_{i+2}(\alpha) \to +\infty$ (or the ends of the spectrum of μ), and as $\alpha \to 0$, $x_i(\alpha) \to x_i^*$, $x_{i+1}(\alpha) \to x_i^*$, and $x'_{i+1}(\alpha) \to$

x_{i+1}^*, $x'_{i+2}(\alpha) \to x_{i+1}^*$. Lemma 5 below asserts that $x_i^* < x_{i+1}^*$ but first it is necessary to prove Lemma 4.

LEMMA 4. *$\partial p(x_i^*, \omega)/\partial\omega$ does not vanish at ω_i^0 or ω_{i+1}^0 but does vanish for some ω_i^* where $\omega_i^0 < \omega_i^* < \omega_{i+1}^0$, $i = 1, \cdots, n-2$.*

Proof. By the mean value theorem for some $\omega_i^*(\alpha) \varepsilon [\omega_i^0, \omega_{i+1}^0]$,

$$\frac{\partial}{\partial\omega}\int_{x_i(\alpha)}^{x_{i+1}(\alpha)} p(x, \omega)\, d\mu(x)\Bigg|_{\omega=\omega_i^*(\alpha)} = \int_{x_i(\alpha)}^{x_{i+1}(\alpha)} \frac{\partial}{\partial\omega} p(x, \omega_i^*(\alpha))\, d\mu(x) = 0$$

for every $\alpha \varepsilon [0, 1]$. As $\alpha \to 0$, $\omega_i^*(\alpha) \to \omega_i^*$; $\partial p(x_i^*, \omega_i^*)/\partial\omega = 0$. Suppose $\omega_i^* = \omega_i^0$. Then, $\partial p(x_i^*, \omega)/\partial\omega > 0$ for $\omega > \omega_i^0$ which implies that $p(x_i^*, \omega_i^0) < p(x_i^*, \omega_{i+1}^0)$. Since $p(x, \omega)$ is continuous in each variable, there exists $\epsilon > 0$ such that $p(x, \omega_i^0) < p(x, \omega_{i+1}^0)$ for all x satisfying $|x - x_i^*| < \epsilon$. But this implies that for sufficiently small α,

$$\int_{x_i(\alpha)}^{x_{i+1}(\alpha)} p(x, \omega_i^0)\, d\mu(x) < \int_{x_i(\alpha)}^{x_{i+1}(\alpha)} p(x, \omega_{i+1}^0)\, d\mu(x),$$

a contradiction of the definition of $x_i(\alpha)$ and $x_{i+1}(\alpha)$. Similarly, $\omega_i^* \neq \omega_{i+1}^0$. Thus, $\omega_i^* \varepsilon (\omega_i^0, \omega_{i+1}^0)$.

LEMMA 5. *$x_i^* < x_{i+1}^*$, $i = 1, \cdots, n-2$.*

Proof. By Lemma 3, $x_i^* \leqq x_{i+1}^*$. Suppose $x_i^* = x_{i+1}^*$. Then, $\partial p(x_i^*, \omega_i^*)/\partial\omega = \partial p(x_i^*, \omega_{i+1}^*)/\partial\omega = 0$, where $\omega_i^* \varepsilon (\omega_i^0, \omega_{i+1}^0)$, $\omega_{i+1}^* \varepsilon (\omega_{i+1}^0, \omega_{i+2}^0)$, which is impossible by assumption.

This lemma can now be utilized to construct decision procedures possessing the property of M.P. For the 2-action problem any monotone procedure (defined by a single number x_1) is unbiased. In the 3-action problem each monotone procedure (x_1, x_2) which satisfies $h_2(\omega_1^0) = \alpha = h_2(\omega_2^0)$ for some $\alpha \varepsilon [0, 1]$ is unbiased. This means the monotone M.P. procedures are a one parameter family since once x_1 is specified as possible, x_2 and α are determined. For $n = 4$ consider $x_1(\alpha_1)$, $x_2(\alpha_1)$ defined by $h_2(\omega_1^0) = \alpha_1 = h_2(\omega_2^0)$ and $x'_2(\alpha_2)$, $x'_3(\alpha_2)$ defined by $h_3(\omega_2^0) = \alpha_2 = h_3(\omega_3^0)$, where α_1 and α_2 are chosen small enough to insure that $x_2(\alpha_1) < x'_2(\alpha_2)$. By Lemma 5 this is possible. Increase α_1 and α_2 until $x_2(\alpha_1) = x'_2(\alpha_2)$. The monotone procedure defined by $(x_1(\alpha_1), x_2(\alpha_1), x'_3(\alpha_2))$ has the property of M.P. Again the monotone M.P. procedures form a one parameter family since any point $y \varepsilon (x_1^*, x_2^*)$ will determine α_1 and α_2 by the condition that $x_2(\alpha_1) = y = x'_2(\alpha_2)$.

For the case of 5 actions the same method of construction is employed and a one parameter family of monotone M.P. decision procedures is designated. Define

$$x_1(\alpha_1),\ x_2(\alpha_1) \text{ by } h_2(\omega_1^0) = \alpha_1 = h_2(\omega_2^0),$$
$$x'_2(\alpha_2),\ x'_3(\alpha_2) \text{ by } h_3(\omega_2^0) = \alpha_2 = h_3(\omega_3^0),$$
$$x''_3(\alpha_3),\ x''_4(\alpha_3) \text{ by } h_4(\omega_3^0) = \alpha_3 = h_4(\omega_4^0),$$

where α_1, α_2, α_3 are chosen so small that $x_2(\alpha_1) < x'_2(\alpha_2)$ and $x'_3(\alpha_2) < x''_3(\alpha_3)$. Increase α_1 and α_3 until $x_2(\alpha_1) = x'_2(\alpha_2)$ and $x'_3(\alpha_2) = x''_3(\alpha_3)$. The monotone procedure $(x_1(\alpha_1), x_2(\alpha_1), x'_3(\alpha_2), x''_4(\alpha_3))$ has the property of maximum probabilities.

The family has only one parameter since the point $y \varepsilon (x_1^*, x_2^*)$ determines α_1, α_2, and α_3 through the relation $x_2(\alpha_1) = y = x_2'(\alpha_2)$. (Note that some values of y in the interval may not be legitimate parameter points. This will happen when the condition $y = x_2'(\alpha_2)$ is satisfied by an α_2 for which $x_3'(\alpha_2) > x_3^*$.)

When $n = 6$, the reader may verify that this method of construction breaks down. The difficulty is that $x_i(\alpha)$ does not have to decrease at the same rate at which $x_{i+1}(\alpha)$ increases. It may not be possible to choose α_2 and α_3 such that $x_3'(\alpha_2) = x_3''(\alpha_3)$ and still have $x_1^* < x_2'(\alpha_2)$ and $x_4''(\alpha_3) < x_4^*$.

For the cases $n = 3, 4$, and 5, note what has been accomplished by introducing the principle of M.P. The statistician, instead of having to choose a procedure from the class of all monotone procedures which is defined by $n - 1$ parameters, has only to choose from a class of procedures defined by only one parameter, those monotone procedures which have the additional property of maximum probabilities.

If the unknown parameter occurs in the density in the form of a translation parameter, that is $p(\xi, \omega) = p(\xi - \omega)$, $d\mu(\xi) = d\xi$, and $p(\cdot)$ is a symmetric function with respect to the origin, then any monotone procedure φ^0 defined by the critical numbers $x_1 < x_2 < \cdots < x_{n-1}$ such that

$$\frac{x_i + x_{i+1}}{2} = \frac{\omega_i^0 + \omega_{i+1}^0}{2} \text{ for } i = 1, 2, \cdots, n-2$$

satisfies the property of M.P. The proof of this statement is straightforward and is omitted.

3. Unbiasedness in the sense of Lehmann—A decision procedure $\varphi(x)$ is said to be unbiased (in the sense of Lehmann or risk unbiased) if

$$E_\theta[L(\omega, \varphi(x))] \geqq E_\theta[L(\theta), \varphi(x))] \tag{11}$$

for all ω and θ, where $E_\theta(\cdot)$ represents the expected value given that the state of nature is θ. By specializing the loss function $L(\omega, a)$, it can be readily verified that this general definition of unbiasedness reduces to some of the classical notions. For a full discussion of the significance of this concept, the reader is referred to [7].

We search in this analysis to discover when unbiased procedures exist within the class of monotone procedures for the case of multiaction problems. An effective method of explicit construction of such procedures would also be desirable. Unfortunately, in general unbiased procedure need not exist. However, Theorem 1 below provides an affirmative answer for a substantial class of loss functions satisfying assumptions (a) and (b).

It should be emphasized that in contrast to the principle of M.P., which also embodies a generalization of the notion of unbiasedness in testing hypotheses, the present extension involves the specific loss functions in a fundamental way.

$$\text{(a)} \qquad L_j(\omega) = L_{ij} \qquad \text{for all } \omega \text{ in } S_i = (\omega_{i-1}^0, \omega_i^0],$$
$$i = 1, \cdots, n, \quad j = 1, \cdots, n.$$

Let $L_{ij} - L_{i+1,j} = a_{i,j}$.

(b) $0 \geqq a_{i1} \geqq a_{i2} \geqq \cdots \geqq a_{ii}$ and $a_{ii} < 0$;

$$a_{i,i+1} \geqq a_{i,i+2} \geqq \cdots a_{i,n} \geqq 0 \text{ and } a_{i,i+1} > 0 \text{ for } i = 1, 2, \cdots, n-1.$$

$$\text{Let } b_{ij} = \begin{cases} -a_{ij} & j = 1, \cdots, i \\ & \qquad\qquad i = 1, \cdots, n-1 \\ a_{ij} & j = i+1, \cdots, n. \end{cases}$$

For $j \leqq$ i, $k \geqq i + 2$, $\qquad i = 1, \cdots, n-1,$

(c) $$\begin{vmatrix} b_{i,j} & b_{i,k} \\ b_{i+1,j} & b_{i+1,k} \end{vmatrix} \geqq 0.$$

Two important examples of decision problems whose loss functions satisfy conditions (b) and (c) are worth noting.

(I) $$L_j(\omega) = c \mid i - j \mid \qquad \text{for } \omega \text{ in } S_i.$$

This case is referred to as the discrete absolute error loss function.

(II) $$L_j(\omega) = \begin{cases} 0 & \omega \varepsilon S_j, \\ c & \omega \not\varepsilon S_j. \end{cases}$$

The second example corresponds to the case where one assigns a constant loss c for any error and zero loss for a correct decision.

The fact that, if it exists, the monotone unbiased procedure is unique lends greater significance to this principle.

Examples I and II above are special cases of loss structures having the form $L_{ij} = f(\mid i - j \mid) = L_{\mid i-j \mid}$. Loss structures of this general pattern possess considerable interest since many practical problems arise in which the incurred losses can be assumed to be proportional to the magnitude of the error and unrelated to the type of error. In the event that $L_{ij} = L_{\mid i-j \mid}$ (we say L_{ij} has a convolution form), condition (b) implies that $L_{\mid i-j \mid}$ is a concave function of $\mid i - j \mid$, i.e., $L_{r+1} \geqq \frac{1}{2}(L_r + L_{r+2})$, $r = 0, 1, \cdots, n - 2$. This is to say the loss increases concavely as the action actually taken diverges from the correct action. That concavity implies condition (b) is also true, so condition (b) is fully equivalent to the concavity of $L_{\mid i-j \mid}$ as a function of $\mid i - j \mid$. Moreover, condition (c) is automatically satisfied if $L_{\mid i-j \mid}$ is concave since $b_{ij} \geqq b_{i+1,j}$ for $j \leqq i$ and $b_{i,k} \leqq b_{i+1,k}$ for $k \geqq i + 2$. Therefore, for this convolution case, the hypotheses of Theorem 1 are equivalent to the statement that $L_{\mid i-j \mid}$ is a concave function of its argument.

It should be noted that condition (c) is not the same as condition (II) of [3]. However, in the important case $L_{ij} = L_{\mid i-j \mid}$, the two conditions are equivalent since the two b_{ij} matrices are identical. Consequently, when the loss function $L_{ij} = L_{\mid i-j \mid}$ is concave, all non-degenerate monotone procedures are admissible.

In particular, the unique unbiased procedure guaranteed by Theorem 1 which is also shown to be non-degenerate (Corollary 4) is necessarily admissible in the case where L_{ij} is of the convolution form.

(The proof of Theorem 2 of [3] is easily seen to apply in the case of loss functions of convolution form satisfying (b) and (c), above.)

The principle theorem concerning unbiased procedures is the following:

THEOREM 1. *If assumptions* (a), (b), *and* (c) *are satisfied, then there exists a unique monotone procedure which is unbiased in the sense of Lehmann.*

To avoid inessential tedious details we assume that $p(x, \omega)$ is strictly Pólya type 2, and μ is a continuous measure whose spectrum is an interval. The analogous results when the assumption on μ is relaxed are immediate.

The proof of Theorem 1 is more elaborate and will be presented in Sec. 4. We dwell in this section on the important special case of (I) where the proofs are considerably simpler and for which some additional results are obtained (Theorem 2).

Proof of Theorem 1 *for the special case* (I). For a monotone procedure $(x_1, \cdots, x_{n-1})$ define

$$A_1(\omega) = c\int_{x_1}^{x_2} p(x, \omega)\, d\mu(x) + 2c\int_{x_2}^{x_3} p(x, \omega)\, d\mu(x) + \cdots + (n-1)c\int_{x_{n-1}}^{\infty} p(x, \omega)\, d\mu(x)$$

$$A_2(\omega) = c\int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) + c\int_{x_2}^{x_3} p(x, \omega)\, d\mu(x) + \cdots + (n-2)c\int_{x_{n-1}}^{\infty} p(x, \omega)\, d\mu(x)$$

$$\vdots$$

$$A_n(\omega) = (n-1)c\int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) + (n-2)c\int_{x_1}^{x_2} p(x, \omega)\, d\mu(x) + \cdots + c\int_{x_{n-2}}^{x_{n-1}} p(x, \omega)\, d\mu(x).$$

For $\omega \,\varepsilon\, S_i$, $i = 1, \cdots, n$, $\rho(\omega, \varphi) = A_i(\omega)$. Define

$$B_i(\omega) = A_i(\omega) - A_{i+1}(\omega), \qquad i = 1, \cdots, n-1.$$

It is immediate that

$$B_i(\omega) = -c\int_{-\infty}^{x_i} p(x, \omega)\, d\mu(x) + c\int_{x_i}^{\infty} p(x, \omega)\, d\mu(x), \qquad i = 1, \cdots, n-1.$$

In order that the monotone procedure be unbiased it is necessary and sufficient that $B_j(\omega) \geqq 0$, $j = 1, \cdots, i-1$; $B_j(\omega) \leqq 0$, $j = i, \cdots, n-1$ for $\omega \,\varepsilon\, S_i$, $i = 1, \cdots, n-1$. Choose the unique $x_1 = x_1^0$ which satisfies $B_1(\omega_1^0) = 0$. Then $B_1(\omega)$ $(\lesseqgtr)$ 0 for all ω $(\lesseqgtr)$ ω_1^0. Since $x_i \geqq x_1^0$ for $i = 2, \cdots, n-1$, $B_i(\omega) < 0$ for $\omega < \omega_1^0$, $i = 2, \cdots, n-1$. Unbiasedness further requires that for $\omega \,\varepsilon\, S_2$ $B_1(\omega) \geqq 0$ and $B_i(\omega) \leqq 0$ for $i = 2, \cdots, n-1$. Determine the unique $x_2 = x_2^0$ such that $B_2(\omega_2^0) = 0$. $x_2^0 > x_1^0$ since $\omega_2^0 > \omega_1^0$, and $B_i(\omega) < 0$ for $\omega \,\varepsilon\, S_2$ and $i =$

$2, \cdots, n-1$. The continuation of this construction will produce the unique monotone unbiased procedure $(x_1^0, \cdots, x_{n-1}^0)$.

For the special loss function under consideration this unique unbiased procedure is uniformly most powerful within the class of all unbiased procedures. This is the substance of the following theorem which is a special case of Theorem 2 of [8]. The proof is included by merit of its simplicity and because it also illustrates on a small scale some of the ideas necessary in carrying out the arguments of Theorem 1.

THEOREM 2. *If* $L_i(\omega) = c \mid i - j \mid$ *for* ω *in* S_j, *then any unbiased procedure* $\varphi = (\varphi_1, \cdots, \varphi_n)$ *is everywhere improved upon by the unique monotone unbiased procedure, except possibly at* $\omega_1^0, \cdots, \omega_{n-1}^0$.

Proof. By definition,

$$B_1(\omega) = A_1(\omega) - A_2(\omega) = -c \int_{-\infty}^{\infty} \varphi_1(x)p(x, \omega)\, d\mu(x) + c \int_{-\infty}^{\infty} (1 - \varphi_1(x))p(x, \omega)\, d\mu(x)$$

$$= c - 2c \int_{-\infty}^{\infty} \varphi_1(x)p(x, \omega)\, d\mu(x),$$

and for $k = 2, \cdots, n-1$,

$$B_k(\omega) = A_k(\omega) - A_{k+1}(\omega) = c - 2c \int_{-\infty}^{\infty} [\varphi_1(x) + \cdots + \varphi_k(x)]p(x, \omega)\, d\mu(x).$$

Consider any other decision procedure φ^* which is not necessarily unbiased. For $k = 1, \cdots, n-1$,

$$B_k^{\varphi}(\omega) - B_k^{\varphi^*}(\omega) = 2c \int_{-\infty}^{\infty} [(\varphi_1^*(x) + \cdots + \varphi_k^*(x)) - (\varphi_1(x) + \cdots + \varphi_k(x))]p(x, \omega)\, d\mu(x).$$

If φ^* is the monotone procedure constructed so that it improves upon φ according to Lemma 4 of [4], then φ^* satisfies

$$B_k^{\varphi}(\omega) - B_k^{\varphi^*}(\omega) \begin{cases} \geqq 0 & \text{for } \omega \leqq \omega_k^0 \\ \leqq 0 & \text{for } \omega \geqq \omega_k^0 \end{cases}$$

for $k = 1, \cdots, n-1$. But $B_k^{\varphi}(\omega_k^0) = 0$. Therefore, $B_k^{\varphi^*}(\omega_k^0) = 0$ which implies that φ^* is unbiased. Since there is only one monotone unbiased procedure, φ^* must be identical with the φ^0 of Theorem 1.

The limiting case of an n-action problem as $n \to +\infty$ is an estimation problem. Suppose that for the problem under consideration the limit is taken in such a manner that as $n \to \infty$, $\omega_1^0 \to -\infty$, $\omega_{n-1}^0 \to +\infty$, $\mid \omega_i^0 - \omega_{i-1}^0 \mid \to 0$, $i = 2, \cdots, n-1$, and $L_n(\omega, i_n(a)) \to c \mid a - \omega \mid$, where $i_n(a)$ is defined by $a \,\varepsilon\, S_{i_n(a)}$. The resulting problem is an estimation problem with absolute error loss function. It is easily verified that the estimate $\delta(x)$, which is the limit of the unique monotone procedures which are unbiased in the sense of Lehmann, is defined by the relation

$$\int_{-\infty}^{x} p(y, \delta(x))\, d\mu(y) = \int_{x}^{\infty} p(y, \delta(x))\, d\mu(y). \tag{12}$$

This, of course, states the well-known fact that the median unbiased estimate of θ is the function $\delta(x)$ which satisfies (12) when x is observed.

4. *Proof of Theorem* 1. For purposes of clarity the proof of the theorem is divided into a series of separate steps. First, we introduce the relevant quantities entering into the analysis. For a procedure $\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_n)$, let

$$A_i^\varphi(\omega) = L_{i1} \int_{-\infty}^{\infty} \varphi_1(x)p(x, \omega)\, d\mu(x) + \cdots + L_{in} \int_{-\infty}^{\infty} \varphi_n(x)p(x, \omega) d\mu(x)$$

for $i = 1, \cdots, n$. When ω ranges over S_i the function $A_i^\varphi(\omega)$ coincides with $\rho(\omega, \varphi)$, the expected risk. Also for $i = 1, 2, \cdots, n - 1$, we define

$$\begin{aligned} B_i^\varphi(\omega) &= A_i^\varphi(\omega) - A_{i+1}^\varphi(\omega) \\ &= a_{i1} \int_{-\infty}^{\infty} \varphi_1(x)p(x, \omega)\, d\mu(x) + \cdots + a_{in} \int_{-\infty}^{\infty} \varphi_n(x)p(x, \omega)\, d\mu(x) \\ &= -b_{i1} \int_{-\infty}^{\infty} \varphi_1(x)p(x, \omega)\, d\mu(x) - \cdots - b_{ii} \int_{-\infty}^{\infty} \varphi_i(x)p(x, \omega)\, d\mu(x) \\ &\quad + b_{i,i+1} \int_{-\infty}^{\infty} \varphi_{i+1}(x)p(x, \omega)\, d\mu(x) + \cdots + b_{in} \int_{-\infty}^{\infty} \varphi_n(x)p(x, \omega)\, d\mu(x). \end{aligned} \tag{13}$$

If a decision procedure φ satisfies the system of inequalities

$$\begin{aligned} B_k^\varphi(\omega) &\geqq 0 \qquad 1 \leqq k \leqq i - 1 \\ B_k^\varphi(\omega) &\leqq 0 \qquad i \leqq k \leqq n - 1 \end{aligned} \qquad \text{and } \omega \text{ in } S_i, \tag{14}$$

then φ is clearly unbiased in the sense of Lehmann. In general, the converse is not valid. However, it is true that for monotone procedures the property of unbiasednesss implies that this system of inequalities is satisfied. The inequalities are fulfilled for a monotone procedure $\varphi = (x_1, x_2, \cdots, x_{n-1})$ if and only if

$$B_i^\varphi(\omega_i^0) = 0, \qquad i = 1, 2, \cdots, n - 1. \tag{15}$$

In fact, the variation diminishing properties of the density $p(x, \omega)$ imply that $B_i^\varphi(\omega) < 0$ for $\omega < \omega_i^0$ and $B_i^\varphi(\omega) > 0$ for $\omega > \omega_i^0$ which in turn are equivalent to the system of inequalities (14). Our problem reduces to the demonstration of the existence and uniqueness of a set of values $x = (x_1, x_2, \cdots, x_{n-1})$ where $x_1 \leqq x_2 \leqq \cdots \leqq x_{n-1}$ which are a solution to the system of non-linear equations:

$$\begin{aligned} B_i^x(\omega_i^0) &= -b_{i1} \int_{-\infty}^{x_1} p(\xi, \omega_i^0)\, d\mu(\xi) - \cdots - b_{ii} \int_{x_{i-1}}^{x_i} p(\xi, \omega_i^0)\, d\mu(\xi) \\ &\quad + b_{i,i+1} \int_{x_i}^{x_{i+1}} p(\xi, \omega_i^0)\, d\mu(\xi) + \cdots + b_{i,n} \int_{x_{n-1}}^{\infty} p(\xi, \omega_i^0)\, d\mu(\xi) = 0. \end{aligned} \tag{16}$$

Turning to this task we start by showing that the mapping $x \to y$ which is defined coordinate-wise by $y_i = B_i^x(\omega_i^0)$, $i = 1, \cdots, n-1$, and which maps the $n - 1$ dimensional simplex of all $n - 1$ tuples $x = (x_1, x_2, \cdots, x_{n-1})$ satisfying $x_1 \leqq x_2 \leqq \cdots \leqq x_{n-1}$ into Euclidean $n - 1$ dimensional space (E^{n-1}) is a one-to-one mapping. Precisely:

LEMMA 6. *The mapping $y_i = B_i^x(\omega_i^0)$, $i = 1, \cdots, n-1$, defined on the set of all monotone procedures by means of the formulas* (16) *with image in E^{n-1} space is a one-to-one transformation.*

Proof (by contradiction). Suppose there exist two different monotone procedures $\varphi \sim x = (x_1, x_2, \cdots, x_{n-1})$ and $\varphi' \sim x' = (x_1', x_2', \cdots, x_{n-1}')$ with the property that $B_i^{x'}(\omega_i^0) - B_i^x(\omega_i^0) = 0$ for $i = 1, \cdots, n-1$. Without loss of generality assume $x_1' \geqq x_1$. $B_i^{x'}(\omega_i^0) - B_i^x(\omega_i^0) = 0$, $i = 1, \cdots, n-1$, yields the system of equations

$$
\begin{aligned}
0 = {} & -(b_{11} + b_{12}) \int_{x_1}^{x_1'} p(x, \omega_1^0)\, d\mu(x) + (b_{12} - b_{13}) \int_{x_2}^{x_2'} p(x, \omega_1^0)\, d\mu(x) \\
& + \cdots + (b_{1,n-1} - b_{1n}) \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_1^0)\, d\mu(x) \\
0 = {} & (b_{22} - b_{21}) \int_{x_1}^{x_1'} p(x, \omega_2^0)\, d\mu(x) - (b_{22} + b_{23}) \int_{x_2}^{x_2'} p(x, \omega_2^0)\, d\mu(x) \\
& + \cdots + (b_{2,n-1} - b_{2,n}) \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_2^0)\, d\mu(x) \\
& \vdots \\
0 = {} & (b_{n-1,2} - b_{n-1,1}) \int_{x_1}^{x_1'} p(x, \omega_{n-1}^0)\, d\mu(x) \\
& + \cdots + (b_{n-1,n-1} - b_{n-1,n-2}) \int_{x_{n-2}}^{x_{n-2}'} p(x, \omega_{n-1}^0)\, d\mu(x) \\
& - (b_{n-1,n-1} + b_{n-1,n}) \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_{n-1}^0)\, d\mu(x).
\end{aligned}
$$

Since $(b_{11} + b_{12}) > (b_{12} - b_{13} + \cdots + (b_{1,n-1} - b_{1,n})$, it follows that there exists a k, $1 < k \leqq n - 1$, such that

$$\int_{x_l}^{x_l'} p(x, \omega_1^0)\, d\mu(x) < \int_{x_k}^{x_k'} p(x, \omega_1^0)\, d\mu(x)$$

for $1 \leqq l < k$. If k is not unique, choose the largest k which satisfies this property. Consider the kth equation. For $1 \leqq l < k$,

$$\int_{x_l}^{x_l'} p(x, \omega_k^0)\, d\mu(x) < \int_{x_k}^{x_k'} p(x, \omega_k^0)\, d\mu(x)$$

by the fundamental change of sign property for strictly Pólya type 2 densities since $x_l \leqq x_k$ and $x_l' < x_k'$. But $(b_{kk} + b_{k,k+1}) \geqq (b_{k2} - b_{k1}) + \cdots + (b_{kk} - b_{k,k-1})$

$+ (b_{k,k+1} - b_{k,k+2}) + \cdots + (b_{k,n-1} - b_{k,n})$. Therefore on examination of the kth equation, if $k < n - 1$, there exists an $h > k$ for which

$$\int_{x_l}^{x'_l} p(x, \omega_k^0)\, d\mu(x) < \int_{x_h}^{x'_h} p(x, \omega_k^0)\, d\mu(x)$$

for $1 \leqq l < h$. If h is not unique, choose the largest h.

Continue this argument until at the last step it has been established that

$$\int_{x_l}^{x'_l} p(x, \omega_{n-1}^0)\, d\mu(x) < \int_{x_{n-1}}^{x'_{n-1}} p(x, \omega_{n-1}^0)\, d\mu(x)$$

for $1 \leqq l < n - 1$. But this contradicts the fact that $B_{n-1}^{x'}(\omega_{n-1}^0) - B_{n-1}^{x}(\omega_{n-1}^0) = 0$ since $(b_{n-1,n-1} + b_{n-1,n}) > (b_{n-1,2} - b_{n-1,1}) + \cdots + (b_{n-1,n-1} - b_{n-1,n-2})$.

COROLLARY 1. *There exists at most one monotone unbiased procedure.*

The proof is immediate. We shall need the following slight extension of Lemma 6.

Corollary 2. *If φ is the monotone procedure $x = (x_1, x_2, \cdots, x_{n-1})$ and $\varphi' \sim x' = (x'_1, x'_2, \cdots, x'_{n-1})$ with $x'_{n-1} \geqq x_{n-1}$ and $B_i^{x'}(\omega_i^0) - B_i^{x}(\omega_i^0) \geqq 0$ for $i = 1, 2, \cdots, n - 1$, then $x'_i = x_i$ for $i = 1, 2, \cdots, n - 1$.*

The proof of Corollary 2 is essentially a paraphrase of that of Lemma 6. We sketch the details. Let k be the first index where $x'_k \geqq x_k$ ($k \leqq n - 1$). By examining the kth relation $B_k^{x'}(\omega_k^0) - B_k^{x}(\omega_k^0) \geqq 0$ as in the proof of the lemma, when $k < n - 1$, we may find a larger index $h > k$ such that for $i < h$,

$$\int_{x_i}^{x'_i} p(\xi, \omega_k^0)\, d\mu(\xi) < \int_{x_h}^{x'_h} p(\xi, \omega_k^0)\, d\mu(\xi).$$

From the variation diminishing properties of $p(\xi, \omega)$ we may conclude that for $i < h$,

$$\int_{x_i}^{x'_i} p(\xi, \omega_h^0)\, d\mu(\xi) < \int_{x_h}^{x'_h} p(\xi, \omega_h^0)\, d\mu(\xi).$$

On continued inspection of the hth relation, we find a larger index until we reach the $(n - 1)^{\text{th}}$ index with the property that

$$\int_{x_i}^{x'_i} p(\xi, \omega_{n-1}^0)\, d\mu(\xi) < \int_{x_{n-1}}^{x'_{n-1}} p(\xi, \omega_{n-1}^0)\, d\mu(\xi), \qquad i = 1, 2, \cdots, n - 2.$$

The last inequality

$$B_{n-1}^{x'}(\omega_{n-1}^0) - B_{n-1}^{x}(\omega_{n-1}^0) \geqq 0$$

is evidently contradicted.

One final extension in the same direction is the following:

COROLLARY 3. *If $\varphi \sim (x_1, x_2, \cdots, x_{n-1})$ and $\varphi' \sim (x'_1, x'_2, \cdots, x'_{k-1}, \gamma, \cdots, \gamma)$*

are two monotone procedures such that $B_i^{\varphi'}(\omega_i^0) - B_i^{\varphi}(\omega_i^0) \geqq 0$ *for* $i = 1, \cdots, k-1$ *and* $\gamma \geqq x_{n-1}$, *then*

$$B_k^{\varphi'}(\omega_k^0) - B_k^{\varphi}(\omega_k^0) \leqq 0.$$

The proof follows the same line of reasoning as the preceding.

In view of Corollary 1 it remains to prove the existence part of Theorem 1. We require the following lemma.

LEMMA 7. *Let the* $2 \times m$ *matrix* (e_{ij}), $i = 1, 2$, $j = 1, \cdots, m$, *consist of non-negative elements, and let* $\lambda_1, \cdots, \lambda_m$ *be non-negative constants. Let condition* (E) *be satisfied:*

$$\text{(E)} \qquad \begin{vmatrix} e_{1j} & e_{1k} \\ e_{2j} & e_{2k} \end{vmatrix} \geqq 0$$

for $1 \leqq j \leqq l$, $l + 2 \leqq k \leqq m$. *If* $0 < e_{11}\lambda_1 + \cdots + e_{1l}\lambda_l \leqq e_{1,l+2}\lambda_{l+2} + \cdots + e_{1,m}\lambda_m$, *then* $e_{21}\lambda_1 + \cdots + e_{2l}\lambda_l \leqq e_{2,l+2}\lambda_{l+2} + \cdots + e_{2m}\lambda_m$.

Proof. By (E), $\sum_{j=1}^{l} (e_{2k}e_{1j} - e_{1k}e_{2j})\lambda_j \geqq 0$ for $k \geqq l+2$. Therefore, $e_{2k} \sum_{j=1}^{l} e_{1j}\lambda_j \geqq e_{1k} \sum_{j=1}^{l} e_{2j}\lambda_j$, and $\sum_{k=l+2}^{m} e_{2k}\lambda_k \cdot \sum_{j=1}^{l} e_{1j}\lambda_j \geqq \sum_{k=l+2}^{m} e_{1k}\lambda_k \cdot \sum_{j=1}^{l} e_{2j}\lambda_j$. For $0 < \sum_{j=1}^{l} e_{1j}\lambda_j \leqq \sum_{k=l+2}^{m} e_{1k}\lambda_k$, $\sum_{k=l+2}^{m} e_{2k}\lambda_k \geqq \sum_{j=1}^{l} e_{2j}\lambda_j$.

Proof of existence. It suffices to show there exists a monotone φ for which $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n-1$. This holds trivially for $n = 2$. Suppose it is true for the case of n actions. The argument is inductive. For $n + 1$ actions and a monotone procedure, let

$$B_i^{\varphi}(\omega) = -b_{i1} \int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) - \cdots - b_{ii} \int_{x_{i-1}}^{x_i} p(x, \omega)\, d\mu(x)$$
$$+ b_{i,i+1} \int_{x_i}^{x_{i+1}} p(x, \omega)\, d\mu(x) + \cdots + b_{i,n+1} \int_{x_n}^{\infty} p(x, \omega)\, d\mu(x)$$

for $i = 1, \cdots, n$.

(1) Choose $x_n = \infty$. The conditions (a), (b), and (c) are fulfilled so by the induction hypothesis there exists a solution $\varphi^{\infty} \sim (x_1^{\infty}, \cdots, x_{n-1}^{\infty}, \infty)$ of the system of equations $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n-1$. For this solution obviously $B_n^{\infty}(\omega_n^0) \leqq 0$.

(2) Choose $x_{n-1} = x_n$. By the induction hypothesis there exists a solution $\varphi^{x^0} \sim (x_1^0, \cdots, x_{n-1}^0 = x^0, x_n^0 = x^0)$ of $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n-1$. Since $B_{n-1}^{x^0}(\omega_{n-1}^0) = 0$, the variation diminishing properties of densities possessing a strict monotone likelihood ratio lead to the conclusion that $B_{n-1}^{x^0}(\omega_n^0) \geqq 0$. If $B_{n-1}^{x^0}(\omega_n^0) = 0$, then it follows that $x^0 = -\infty$ which in turn implies that $B_n^{x^0}(\omega_n^0) > 0$. On the other hand, if $B_{n-1}^{x^0}(\omega_n^0) > 0$, let $l = n-1$, $m = n$, $e_{1j} = b_{n-1,j}$ for $j = 1, \cdots, n-1$, $e_{1n} = b_{n-1,n+1}$, $e_{2j} = b_{n,j}$ for $j = 1, \cdots, n-1$ and $e_{2n} = b_{n,n+1}$ in Lemma 7. Then, by Lemma 7 $B_{n-1}^{x^0}(\omega_n^0) > 0$ implies $B_n^{x^0}(\omega_n^0) \geqq 0$.

It has been shown thus far that there exists a strategy $(x_1^{\infty}, \cdots, x_{n-1}^{\infty}, \infty)$ such that $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n-1$, and $B_n^{\varphi}(\omega_n^0) \leqq 0$ and a strategy $(x_1^0, \cdots, x_{n-1}^0 = x^0, x_n^0 = x^0)$ such that $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n-1$ and $B_n^{\varphi}(\omega_n^0) \geqq 0$. If it can be shown that for every x_n satisfying $x^0 < x_n < \infty$ there exists a solution $(x_1, \cdots, x_{n-1})$ to $B_i^\varphi(\omega_i^0) = 0$ $i = 1, \cdots, n-1$, then by continuity a solution exists satisfying $B_i^\varphi(\omega_i^0) = 0$, $i = 1, \cdots, n$; the continuity of the solution as a function of x_n being a simple consequence of Lemma 6.

The proof that for every z, $x^0 < z < \infty$, there exist $(x_1(z), \cdots, x_{n-1}(z))$ such that $\varphi \sim (x_1(z), \cdots, x_{n-1}(z), z)$ satisfies $B_i^\varphi(\omega_i^0) = 0$, $i = 1, \cdots, n-1$, proceeds in a stepwise manner.

(3) Let $x_1 \leqq x_2 = \cdots = x_{n-1} = z$.

(a) Choose $x_1 = z$. Since $b_{12} \geqq b_{13} \geqq \cdots \geqq b_{1,n+1}$,

$$-b_{11} \int_{-\infty}^{x_1^0} p(x, \omega_1^0)\, d\mu(x) + b_{1,n+1} \int_{x_1^0}^{\infty} p(x, \omega_1^0)\, d\mu(x) \leqq 0$$

which implies

$$-b_{11} \int_{-\infty}^{z} p(x, \omega_1^0)\, d\mu(x) + b_{1,n+1} \int_{z}^{\infty} p(x, \omega_1^0)\, d\mu(x) \leqq 0$$

since $x_1^0 \leqq x_n^0 < z$.

(b) Choose $x_1 = -\infty$.

$$b_{12} \int_{-\infty}^{z} p(x, \omega_1^0)\, d\mu(x) + b_{1,n+1} \int_{z}^{\infty} p(x, \omega_1^0)\, d\mu(x) \geqq 0.$$

(c) Thus by continuity there must exist an $x_1^1 = x_1^1(z)$ which satisfies

$$-b_{11} \int_{-\infty}^{x_1^1} p(x, \omega_1^0)\, d\mu(x) + b_{12} \int_{x_1^1}^{z} p(x, \omega_1^0)\, d\mu(x) + b_{1,n+1} \int_{z}^{\infty} p(x, \omega_1^0)\, d\mu(x) = 0.$$

(4) Let $x_1 \leqq x_2 \leqq x_3 = \cdots = x_{n-1} = z$. Consider the two expressions

$$\begin{aligned} c_1(\omega; x_1, x_2) = &-b_{11} \int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) + b_{12} \int_{x_1}^{x_2} p(x, \omega)\, d\mu(x) \\ &+ b_{13} \int_{x_2}^{z} p(x, \omega)\, d\mu(x) + b_{1,n+1} \int_{z}^{\infty} p(x, \omega)\, d\mu(x), \\ c_2(\omega; x_1, x_2) = &-b_{21} \int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) - b_{22} \int_{x_1}^{x_2} p(x, \omega)\, d\mu(x) \\ &+ b_{23} \int_{x_2}^{z} p(x, \omega)\, d\mu(x) + b_{2,n+1} \int_{z}^{\infty} p(x, \omega)\, d\mu(x). \end{aligned}$$

Of course $c_j(\omega, x_1, x_2) = B_j^\varphi(\omega)$, $j = 1, 2$, for the special procedure $\varphi \sim (x_1, x_2, z, \cdots z)$. Our immediate object now is to show that x_1 and x_2 exist satisfying $(x_1 \leqq x_2 \leqq z)$ such that $c_1(\omega_1^0; x_1, x_2) = 0$ and $c_2(\omega_2^0; x_1, x_2) = 0$.

(a) Choose $x_2 = z$. By (3) above there exists an $x_1'(z)$ for which

$$c_1(\omega_1^0; x_1'(z), z) = 0.$$

We assert that $c_2(\omega_2^0; x_1'(z), z) \leqq 0$. Comparing for $i = 1, 2$ $B_i^{\varphi^0}(\omega_i^0)$ and $B_i^\varphi(\omega_i^0)$ where $\varphi^0 \sim (x_1^0, x_2^0, \cdots, x_{n-1}^0 = x^0, x_n^0 = x^0)$ of (2) above and

$\varphi \sim (x_1'(z), z, z, \cdots, z)$ with $z > x^0$, we see the conditions of Corollary 3 are met and therefore we may conclude $c_2(\omega_2^0, x_1'(z), z) \leqq 0$ as stated.

(b) Choose $x_1 = x_2$; then $c_1(\omega_1^0; -\infty, -\infty) \geqq 0$ and $c_1(\omega_1^0; z, z) \leqq 0$ by (3a). Thus there exists a $u = x_1 = x_2$ such that $c_1(\omega_1^0; u, u) = 0$ which implies $c_1(\omega_2^0; u, u) \geqq 0$ If $c_1(\omega_2^0; u, u) = 0$, then $u = -\infty$ which in turn implies $c_2(\omega_2^0; u, u) \geqq 0$. If in the other circumstance $c_1(\omega_2^0; u, u) > 0$, then by Lemma 7 we infer that $c_2(\omega_2^0; u, u) \geqq 0$.

(c) We next prove that there exists an $x_1^* = x_1^*(y)$ such that $c_1(\omega_1^0; x_1^*, y) = 0$ for every $u < y < z$. (This is like the larger problem we are trying to solve for the special case when $n = 2$. The quantity z plays the role of ∞ and u adopts the role of z.) When $x_1 = y$, $c_1(\omega_1^0; y, y) < 0$ because $c_1(\omega_1^0; u, u) = 0$ and $y > u$. Obviously $c_1(\omega_1^0; -\infty, y) > 0$. By continuity there exists an x_1^* such that $c_1(\omega_1^0; x_1^*, y) = 0$.

Since $c_1(\omega_1^0; x_1, y) = 0$ has a solution x_1^* for every y in the interval $[u, z]$ and $c_2(\omega_2^0; x_1'(z), z) \leqq 0$, $c_2(\omega_2^0; u, u) \geqq 0$, by continuity there must exist an $x_2^2(z) = y \in [u, z]$ and $x_1^2(z)$ such that $c_1(\omega_1^0; x_1^2, x_2^2) = c_2(\omega_2^0; x_1^2, x_2^2) = 0$.

(5) Let $x_1 \leqq x_2 \leqq x_3 \leqq x_4 = \cdots = z$. Consider the three expressions

$$\begin{aligned} D_i(\omega; x_1, x_2, x_3) = &- \sum_{j=1}^{i} b_{ij} \int_{x_{j-1}}^{x_j} p(x, \omega)\, d\mu(x) \\ &+ \sum_{j=i+1}^{3} b_{ij} \int_{x_{j-1}}^{x_j} p(x, \omega)\, d\mu(x) + b_{i4} \int_{x_3}^{z} p(x, \omega)\, d\mu(x) \\ &+ b_{i,n+1} \int_{z}^{\infty} p(x, \omega)\, d\mu(x), \end{aligned}$$

$i = 1, 2, 3$, where $x_0 = -\infty$. Of course $D_i(\omega; x_1, x_2, x_3) = B_i^\varphi(\omega)$ where $\varphi \sim (x_1, x_2, x_3, z, z, \cdots, z)$. The next step is to try to solve $D_i(\omega_i^0; x_1, x_2, x_3) = 0$, $i = 1, 2, 3$.

(a) Choose $x_3 = z$. By (4) above there exists a couple $(x_1^2(z), x_2^2(z))$ such that $D_1(\omega_1^0; x_1^2(z), x_2^2(z), z) = D_2(\omega_2^0; x_1^2(z), x_2^2(z), z) = 0$. Corollary 3 may be applied and we find that on comparison with the relations $B_i^{\varphi^0}(\omega_i) = 0$, $i = 1, 2, 3$, for $\varphi^0 \sim (x_1^0, x_2^0, \cdots, x_n^0)$ of (2), $D_3(\omega_3^0; x_1^2(z), x_2^2(z), z) \leqq 0$.

(b) Choose $x_2 = x_3$. By (4) there exists a solution $(\bar{x}_1(w), w)$ where $x_2 = x_3 = w$ to the equations $D_1(\omega_1^0; x_1, x_2, x_2) = 0$, $D_2(\omega_2^0; x_1, x_2, x_2) = 0$. $D_3(\omega_3^0; \bar{x}_1, w, w) \geqq 0$ is a consequence of Lemma 7.

(c) There exists a couple $(x_1^{**}(y), x_2^{**}(y))$ such that

$$D_1(\omega_1^0; x_1^{**}(y), x_2^{**}(y), y) = D_2(\omega_2^0; x_1^{**}(y), x_2^{**}(y), y) = 0$$

for every $y \in [w, z]$.

The proof of this step requires a repetition of the previous arguments as carried out for the function c_i with y taking the role of ∞. To this end, we establish

(c.1) Choose $x_1 = x_2$. For $x_1 = -\infty$, $D_1(\omega_1^0; -\infty, -\infty, y) \geqq 0$. For $x_1 = y$, $D_1(\omega_1^0; y, y, y) \leqq 0$ since $D_1(\omega_1^0; \bar{x}_1(w), w, w) = 0$ implies $D_1(\omega_1^0; w, w, w) \leqq 0$ and $y > w$. Therefore, there exists $x_1 = v$ such that $D_1(\omega_1^0; v, v, y) = 0$. It can be shown by applying Lemma 7 that $D_2(\omega_2^0; v, v, y) \geqq 0$.

(c.2) Choose $x_2 = y$. $D_1(\omega_1^0\,;\,-\infty, y, y) \geqq 0$ and $D_1(\omega_1^0\,;\,y, y, y) \leqq 0$. Thus, there exists $x_1(y)$ such that $D_1(\omega_1^0\,;\,x_1(y), y, y) = 0$. $D_2(\omega_2^0\,;\,x_1(y), y, y) \leqq 0$. The last inequality may be deduced from Corollary 3 by comparing the procedures $\varphi' \sim (x_1(y), y, y, z, z, \cdots, z)$ and $\varphi \sim (\bar{x}_1(w), w, w, z, z, \cdots, z)$.

In fact, suppose the inequality $D_2(\omega_2^0\,;\,x_1(y), y, y) \leqq 0$ is violated. Consider the solution $(\bar{x}_1(w), w, w)$ to the system of equations

$$D_1(\omega_1^0\,;\,x_1, x_2, x_2) = D_2(\omega_2^0\,;\,x_1, x_2, x_2) = 0.$$

$$\begin{aligned} &D_1(\omega_1^0\,;\,x_1(y), y, y) - D_1(\omega_1^0\,;\,\bar{x}_1(w), w, w) \\ &\quad = -(b_{11} + b_{12}) \int_{\bar{x}_1(w)}^{x_1(y)} p(x, \omega_1^0)\, d\mu(x) + (b_{12} - b_{14}) \int_w^y p(x, \omega_1^0)\, d\mu(x) = 0, \end{aligned} \tag{17}$$

$$\begin{aligned} &D_2(\omega_2^0\,;\,x_1(y), y, y) - D_2(\omega_2^0\,;\,\bar{x}_1(w), w, w) \\ &\quad = (b_{22} - b_{21}) \int_{\bar{x}_1(w)}^{x_1(y)} p(x, \omega_2^0)\, d\mu(x) - (b_{22} + b_{24}) \int_w^y p(x, \omega_2^0)\, d\mu(x) > 0. \end{aligned} \tag{18}$$

Eq. (17) implies $\int_w^y p(x, \omega_1^0)\, d\mu(x) > \int_{\bar{x}_1(w)}^{x_1(y)} p(x, \omega_1^0)\, d\mu(x)$, but this contradicts (18). Therefore, $D_2(\omega_2^0\,;\,x_1(y), y, y) \leqq 0$.

(c.3) For every $x_2 \in [v, y]$, $D_1(\omega_1^0\,;\,x_1, x_2, y) = 0$ has a solution. By continuity, then, there exists an $x_1^{**}(y)$, $x_2^{**}(y)$ such that $D_1(\omega_1^0\,;\,x_1^{**}(y), x_2^{**}(y), y) = D_2(\omega_2^0\,;\,x_1^{**}(y), x_2^{**}(y), y) = 0$.

(a), (b), and (c) of (5) show that there exists a 3-tuple $(x_1^3(z), x_2^3(z), x_3^3(z))$ which satisfies $D_i(\omega_i^0\,;\,x_1^3, x_2^3, x_3^3) = 0, i = 1, 2, 3$.

The steps for $x_1 \leqq x_2 \leqq x_3 \leqq x_4 \leqq x_5 = \cdots = z$ utilize the same principles as those employed above. The general pattern should now be clear to the reader.

The next step would consider the four functions $E_i(\omega;\,x_1, x_2, x_3, x_4) = B_i^\varphi(\omega)$, $i = 1, \cdots, 4$ where $\varphi \sim (x_1, x_2, x_3, x_4, z, z, \cdots, z)$. It is necessary to show that $E_i(\omega_i^0) = 0$, $i = 1, 2, 3, 4$, have a solution in x_1, x_2, x_3, and x_4. This entails repeating the entire preceding argument for the case of one, two, and three functions in each case using a suitable comparison monotone procedure. We sketch the argument. Setting $x_4 = z$ we obtain by (5) that there exists a tuple $(x_1(z), x_2(z), x_3(z), z)$ for which $E_i(\omega_i^0\,;\,x_1(z), x_2(z), x_3(z), z) = 0$ for $i = 1, 2, 3$. Corollary 3 may be applied by using the second procedure $\varphi^0 \sim (x_1^0, x_2^0, \cdots, x_n^0)$ to show that $E_4(\omega_4^0\,;\,x_1(z), x_2(z), x_3(z), z) \leqq 0$. Next put $x_3 = x_4 = t < z$ and again by (5) we obtain a tuple $(x_1(t), x_2(t), t, t)$ for which $E_i(\omega_i^0\,;\,x_1(t), x_2(t), t, t) = 0$ for $i = 1, 2, 3$. According to Lemma 7, $E_4(\omega_4^0\,;\,x_1(t), x_2(t), t, t) \geqq 0$. Given y, $t < y < z$, it would be enough to construct a solution to $E_i(\omega_i^0\,;\,x_1, x_2, x_3, y) = 0$, $i = 1, 2, 3$, for then by continuity there would exist a solution to $E_i(\omega_i^0) = 0$, $i = 1, 2, 3, 4$. The analysis of $E_i(\omega_i^0\,;\,x_1, x_2, x_3, y)$, $i = 1, 2, 3$, is similar to the arguments of (5) this time using the comparison procedure

$$\varphi \sim (x_1(t), x_2(t), t, t, z, z, \cdots, z)$$

as $\varphi \sim (\bar{x}_1(w), w, w, z, \cdots, z)$ was used in (5). For the final step we repeat this sequence of arguments $n - 1$ times. This completes the proof of Theorem 1.

COROLLARY 4. *The unique monotone unbiased procedure defined by Theorem 1 is non-degenerate.*

Remark. Since the density $p(x, \omega)$ is assumed to have a strict monotone likelihood ratio, the set $\sigma_\omega = \{x \mid p(x, \omega) > 0\}$ is independent of ω [4]. The concept of an interval (x_i, x_{i+1}) being degenerate should therefore be understood as taken with respect to $d\mu(x)$.

Proof. Suppose the unique unbiased procedure $x^0 = (x_0, x_1, x_2, \cdots, x_{n-1}, x_n)$ where $x_0 = -\infty$ and $x_n = +\infty$ possesses a degenerate interval. We shall prove that this assumption leads to an absurdity. First, observe that (x_0, x_1) must be non-degenerate. Otherwise, let j_0 be such that (x_{j_0}, x_{j_0+1}) is the first non-degenerate interval and $j_0 \geqq 1$. By condition (b), $B_{j_0-1}^{x^0}(\omega_{j_0-1}^0) > 0$, which contradicts the definition of x^0. Now let i_0 be the earliest interval where (x_{i_0}, x_{i_0+1}) is degenerate. Therefore by what has been established $i_0 \geqq 1$ and also $i_0 < n - 1$ for in the contrary case $B_{n-1}^{x^0}(\omega_{n-1}^0)$ would be negative, Let k_0 denote the smallest index larger than i_0 for which (x_{k_0}, x_{k_0+1}) is non-degenerate. A value of k must exist, for otherwise $B_{i_0}^{x^0}(\omega_{i_0}^0) < 0$.

The strict monotone likelihood ratio possessed by $p(x, \omega)$ implies that

$$(*)\qquad \int_{x_j}^{x_{j+1}} p(\xi, \omega_i^0)\, d\mu(\xi) \int_{x_r}^{x_{r+1}} p(\xi, \omega_k^0)\, d\mu(\xi) \geqq \int_{x_j}^{x_{j+1}} p(\xi, \omega_k^0)\, d\mu(\xi) \int_{x_r}^{x_{r+1}} p(\xi, \omega_i^0)\, d\mu(\xi)$$

for every $j < i_0$ and $r \geqq k_0$ with strict inequality valid for $j = i_0 - 1$ and $r = k_0$. Equation (*) in conjunction with conditions (b) and (c) and $B_{i_0}^{x^0}(\omega_{i_0}^0) = 0$ readily leads to the result

$$B_{k_0}^{x^0}(\omega_{k_0}^0) > 0,$$

which is impossible. This completes the proof.

In any special case this construction is considerably more facile than the general proof shows. We carry this out for the special case whose loss function is (II) of the preceding section. For any prescribed $x_{i-1} < x_i$ a value $x_{i+1}(x_{i+1} > x_i)$ is determined recursively, whenever possible, by

$$(19)\qquad \int_{x_{i-1}}^{x_i} p(\xi, \omega_i^0)\, d\mu(\xi) = \int_{x_i}^{x_{i+1}} p(\xi, \omega_i^0)\, d\mu(\xi)$$

for $i = 1, 2, \cdots, n - 1$ where $x_0 = -\infty$. For x_1 sufficiently near $-\infty$, it is possible to solve (19) for each x_i such that $x_i > x_{i-1}$ and each is near $-\infty$. Allowing x_1 to increase, we observe that each x_i increases; and ultimately for $x_1 < \infty$, x_n reaches ∞. Let x_i^* be the solution of (19) where $x_n^* = +\infty$. The procedure $\varphi^* \sim (x_1^*, x_2^*, \cdots, x_{n-1}^*)$ is the unique monotone unbiased procedure for the case where

$$L_i(\omega) = \begin{cases} c & \omega \varepsilon S_i \\ 0 & \omega \varepsilon S_i \end{cases}.$$

REFERENCES

[1] SAMUEL KARLIN, "Decision theory of Pólya type distributions. Case of two actions I," *Proceedings of the Third Berkeley Symposium on Probability and Statistics*, Vol. 1, University of California Press, 1955, pp. 115–128.

[2] SAMUEL KARLIN, "Pólya type II" "Pólya type distributions, II," *Ann. Math. Stat.*, vol. 28 (1957), pp. 281–308.

[3] SAMUEL KARLIN, "Pólya type III. Admissibility for multi-action problems," *Ann. Math. Stat.*, vol. 28 (1957), pp. 839–860.

[4] SAMUEL KARLIN and HERMAN RUBIN, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, vol. 27 (1956), pp. 272–300.

[5] DAVID BLACKWELL and M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, 1954.

[6] L. SAVAGE, *The Foundations of Statistics*, John Wiley and Sons, 1954.

[7] E. L. LEHMANN, "A general concept of unbiasedness," *Ann. Math. Stat.*, vol. 22 (1951), pp. 587–592.

[8] E. L. LEHMANN, "A theory of some multiple decision problems, I," *Ann. Math. Stat.*, vol. 28 (1957), pp. 1–25.

THE USE OF GROUP DIVISIBLE DESIGNS FOR CONFOUNDED ASYMMETRICAL FACTORIAL ARRANGEMENTS[1]

By Marvin Zelen

National Bureau of Standards

1. Introduction and summary. A factorial experiment involving m factors such that the ith factor has m_i levels is termed an asymmetrical factorial design. If the number of levels is equal to one another the experiment is termed a symmetric factorial experiment. When the block size of the experiment permits only a sub-set of the factorial combinations to be assigned to the experimental units within a block, resort is made to the theory of confounding. With respect to symmetric factorial designs, the theory of confounding has been highly developed by Bose [1], Bose and Kishen [4], and Fisher [11], [12]. An excellent summary of the results of this research appears in Kempthorne [13]. However, these researches are closely related to Galois field theory resulting in (i) only symmetric factorial designs being incorporated into the current theory of confounding; (ii) the common level must be a prime (or power of a prime) number; and (iii) the block size must be a multiple of this prime number.

The theory of confounding for asymmetric designs has not been developed to any great degree. Examples of asymmetric designs can be found in Yates [19], Cochran and Cox [9], Li [15], and Kempthorne [13]. Nair and Rao [16] have given the statistical analysis of a class of asymmetrical two-factor designs in considerable detail.

Kramer and Bradley [14] discuss the application of group divisible designs to asymmetrical factorial experiments, however their paper is mainly confined to the two-factor case and its intra-block analysis.[2] It is the purpose of this paper, which was done independently of their work, to outline the general theory for using the group divisible incomplete block designs for asymmetrical factorial experiments.

The use of incomplete block designs for asymmetric factorial experiments results in (i) no restriction that the levels must be a prime (or power of a prime) number, (ii) no restriction with respect to the dependence of the block size on the type of level, and (iii) unlike the previous referenced works on asymmetric factorial designs, the resulting analysis is simple, does not increase in difficulty with an increasing number of factors, and "automatically adjusts" for the effects of partial confounding.

Received January 10, 1957; revised June 18, 1957; revised November 1, 1957.

[1] This paper is an extension of results presented at the Annual Meeting of the American Statistical Society, September, 1954 (cf. [22]).

[2] *Note added in proof:* The Editor has pointed out that the paper by K. R. Nair, "A note on group divisible incomplete block designs", *Calcutta Statistical Association Bulletin*, Vol. 5, No. 17, (1953), pp. 30–35, together with Nair and Rao [16] essentially contains the results for the intra-block analysis of the two-factor asymmetrical designs.

Section 2 states three useful lemmas, Section 3 contains the main results of this paper, and Section 4 outlines the recovery of inter-block information.

2. Some useful lemmas.

We state here three lemmas which will be referred to in later sections. Since the proofs are trivial they are omitted.

Let $X' = (X_1, X_2, \cdots, X_n)$ have a multivariate normal distribution such that

$$E(X') = m' = (m_1, m_2, \cdots, m_n),$$

$$E[(X - m)(X - m)'] = M\sigma^2.$$

LEMMA 2.1. *The expected value of the quadratic form $X'AX$ is*

$$E(X'AX) = m'Am + \sigma^2 \text{ trace } (AM).$$

LEMMA 2.2. *If $M^2 = \lambda M$ (λ a scalar), then the quadratic form*

$$\frac{(X - m)'(X - m)}{\lambda}$$

follows a $\sigma^2\chi^2$ distribution with r degrees of freedom where $r \leqq n$ is the rank of M.

LEMMA 2.3. *Define the direct-product of two square matrices A and B of dimensions m and n respectively by*

$$(A*B) = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{bmatrix}.$$

If $A^2 = \alpha A$ and $B^2 = \beta B$ (α and β are scalars), then $(A * B)^2 = \alpha\beta(A * B)$. In general, given p matrices $A, B, C, \cdots$ such that $A^2 = \alpha A$, $B^2 = \beta A$, $C^2 = \gamma C$, $\cdots$ we have $(A * B * C * \cdots)^2 = (\alpha\beta\gamma \cdots)(A * B * C * \cdots)$.

3. Analysis of group divisible designs used as asymmetrical factorials.

3.1. Estimation. The group divisible designs are partially balanced incomplete block designs with two associate classes. These were first discussed extensively by Bose and Connor [3] and Bose and Shimamato [5]. A large catalogue of such experiment plans giving full details of the analysis can be found in Bose, Clatworthy, and Shrikhande [2]. Designs with block size $k = 2$ have been enumerated by Clatworthy [7]. Bose, Shrikhande, and Battacharya [6], and Clatworthy [8] give methods for constructing group divisible designs.

Briefly group divisible designs can be characterized by having b blocks with

k experimental units such that each of the $v = mn$ treatments is replicated r times. The $v = mn$ treatments can be divided into m groups of n treatments each, where any two treatments in the same group are 1st associates and two treatments in different groups are 2nd associates. With respect to any treatment, there will be $(n - 1)$ 1st associates and $n(m - 1)$ 2nd associates.

Consider a factorial experiment with $(g + h)$ factors $A_1, A_2, \cdots, A_g, B_1, \cdots, B_h$ such that the number of levels associated with A_s is m_s for $s = 1, \cdots, g$ and the number of levels associated with B_r is n_r $r = 1, 2, \cdots, h$. Furthermore, let these levels be such that $m = \prod_{s=1}^{g} m_s$ and $n = \prod_{r=1}^{h} n_r$. Then one can use the group divisible designs for a $\prod_{s=1}^{g} m_s \times \prod_{r=1}^{h} n_r$ factorial design by arranging the $v = mn$ treatments in an $n \times m$ array and assigning the m factorial combinations among the A factors to the columns (groups) and the n factorial combinations among the B factors to the rows.

Let the measurement of the uth treatment combination $(u = 1, 2, \cdots, v)$ measured in the zth block be denoted by y_{uz} and let the underlying mathematical model be

$$y_{uz} = m + t_u + b_z + \epsilon_{uz}, \tag{3.1}$$

where m is a constant common to all measurements, t_u is the effect of the uth treatment combination, b_z is the constant associated with the zth block

$$(z = 1, 2, \cdots, b),$$

and $\{\epsilon_{uz}\}$ is a sequence of uncorrelated random variables having a zero mean and (unknown) variance σ^2. For making all tests of significance, we shall further assume that the $\{\epsilon_{uz}\}$ follow a normal distribution.

Due to the factorial nature of the experiment, a treatment combination t_z can be written as

$$\begin{aligned} t_z = {} & \sum_{s=1}^{g} (a_s)_{i_s} + \sum_{q=1}^{h} (b_q)_{j_q} \\ & + \sum_{t=2}^{g} \sum_{s=1}^{t} (a_{st})_{i_{st}} + \sum_{r=2}^{h} \sum_{q=1}^{r} (b_{qr})_{j_{qr}} \\ & + \sum_{q=1}^{h} \sum_{s=1}^{g} (a_s b_q)_{i_s j_q} + \cdots + (a_{12\cdots g} b_{12\cdots h})_{i_{12\cdots g} j_{12\cdots h}}. \end{aligned} \tag{3.2}$$

The $(a_s)_{i_s}$ are constants associated with the main effect of A_s at level i_s; the $(a_{st})_{i_{st}}$ are constants associated with the two factor interaction between A_s and A_t at levels i_s and i_t, etc. Similar interpretations hold for the constants associated with the main effects and interactions of the B factors, and also for the constants associated with the interactions composed of both A and B factors. It is well known that these parameters are not all linearly independent and satisfy the following relations:

$$(3.3)\quad \begin{cases} \sum_{i_s=1}^{m_s} (a_s)_{i_s} = 0, & s = 1, 2, \cdots, g, \\ \sum_{j_q=1}^{n_q} (b_q)_{j_q} = 0, & q = 1, 2, \cdots, h, \\ \sum_{i_\alpha=1}^{m_\alpha} (a_{st})_{i_{st}} = 0, & \alpha = s, t;\ \ s < t = 1, 2, \cdots, g, \\ \sum_{j_\beta=1}^{n_\beta} (b_{qr})_{j_{qr}} = 0, & \beta = q, r;\ \ q < r = 1, 2, \cdots, h, \\ \quad \vdots & \quad \vdots \\ \sum_{i_\alpha=1}^{m_\alpha} (a_{12\cdots g} b_{12\cdots h})_{i_{12}\cdots{}_g j_{12}\cdots{}_h} & \\ = \sum_{j_\beta=1}^{n_\beta} (a_{12\cdots g} b_{12\cdots h})_{i_{12}\cdots{}_g j_{12}\cdots{}_h} = 0, & \alpha = 1, 2, \cdots g;\ \ \beta = 1, 2, \cdots h. \end{cases}$$

If the adjusted treatment total for the uth treatment is defined by

$$Q_u = (u\text{th treatment total}) - \begin{pmatrix}\text{sum of the block averages in} \\ \text{which the } u\text{th treatment occurs}\end{pmatrix},$$

then the treatment estimates can conveniently be written as

$$(3.4)\qquad \hat{t}_u = \frac{1}{r(k-1)}\,[kQ_u + c_1 S_1(Q_u) + c_2 S_2(Q_u)].$$

Here $S_1(Q_u)$ and $S_2(Q_u)$ are the sum of the adjusted treatment totals for the 1st and 2nd associates with respect to treatment u, and c_1, c_2 are constants calculated from the design parameters. (All catalogues of group divisible designs [2], [5], [7], [8], give numerical values of c_1 and c_2).

Since these estimates satisfy the restraint $\sum_{u=1}^{v} \hat{t}_u = 0$, the variance of a treatment estimate can be written as

$$(3.5)\qquad \operatorname{Var} \hat{t}_u = \left[\frac{vk - [k + (n-1)c_1 + n(m-1)c_2]}{r(k-1)v}\right]\sigma^2,$$

and the covariance between treatments which are (say) sth associates ($s = 1, 2,$) is

$$(3.6)\qquad \operatorname{Cov}(\hat{t}_i, \hat{t}_j) = \left[\frac{vc_s - [k + (n-1)c_1 + n(m-1)c_2]}{r(k-1)v}\right]\sigma^2$$

for $s = 1, 2$.

Let (say) $A_1, A_2, \cdots, A_p$ ($p \leqq g$) and $B_1, B_2, \cdots, B_q$ ($q \leqq h$) be a selection of the A and B factors and let them be associated with the particular levels $i = (i_1, i_2, \cdots, i_p)$ and $j = (j_1, j_2, \cdots, j_q)$ respectively. We define an S-function associated with these particular factors and levels by

$$S[1, 2, \cdots, p;\ 1, 2, \cdots, q \mid i, j] = \frac{\prod_{i=1}^{p} m_i \prod_{j=1}^{q} n_j}{v} \sum_u{}' \hat{t}_u, \tag{3.7}$$

where the summation $\sum_u'$ refers to the sum over all treatment estimates which have the same levels $i_1, i_2, \cdots, i_p; j_1, j_2, \cdots, j_q$ with respect to the factors $A_1, A_2, \cdots, A_p; B_1, B_2, \cdots, B_q$. If an S-function contains no A factors, we shall denote this by $S[0; 1, 2, \cdots, q \mid j]$ with a similar notation for the absence of B factors. (Note that these S-functions are simply the cell averages in any $(p + q)$ way table associated with these factors). Then the expected value of (3.7) is

$$\begin{aligned} E\{S[1, 2, \cdots, p;\ 1, 2, \cdots, q \mid i, j]\} &= \sum_{s=1}^{p} (a_s)_{i_s} + \sum_{r=1}^{q} (b_r)_{j_r} \\ &+ \sum_{t=2}^{p} \sum_{s=1}^{t} (a_{st})_{i_{st}} + \sum_{s=2}^{q} \sum_{r=1}^{s} (b_{rs})_{j_{rs}} + \cdots + (a_{12\cdots p} b_{12\cdots q})_{i_{12}\cdots{}_p j_{12}\cdots{}_q}, \end{aligned} \tag{3.8}$$

where the summations refer to the particular factors A_i $(i = 1, 2, \cdots, p)$, B_j $(j = 1, 2, \cdots, q)$ and the levels $i_{st}\ldots j_{rs}\ldots$ refer only to $i = (i_1, i_2, \cdots, i_p)$ and $j = (j_1, j_2, \cdots, j_q)$. There will be only $(v - 1)$ linearly independent treatment estimates and since the relations (3.3) imply that there exist $(v - 1)$ linearly independent factorial constants, the condition of unbiasedness is sufficient to insure unique estimates of the factorial constants. Therefore the estimates of the main effects and interaction parameters are given by

$$\left\{ \begin{aligned} &(\hat{a}_s)_{i_s} = S[s; 0 \mid i_s], \\ &(\hat{b}_q)_{j_q} = S[0; q \mid j_q], \\ &(\hat{a}_{st})_{i_{st}} = S[s, t; 0 \mid i_s, i_t] - \{S[s; 0 \mid i_s] + S[t; 0 \mid i_t]\}, \\ &(\hat{b}_{qu})_{j_{qu}} = S[0; q, u \mid j_q, j_u] - \{S[0; q \mid j_q] + S[0; u \mid j_u]\}, \\ &(a_s\hat{b}_q)_{i_s j_q} = S[s; q \mid i_s, j_q] - \{S[s; 0 \mid i_s] + S[0; q \mid j_q]\}, \\ &\qquad \vdots \qquad\qquad\qquad\qquad \vdots \\ &(a_{12\cdots g}\hat{b}_{12\cdots h})_{i_{12}\cdots{}_g j_{12}\cdots{}_h} \\ &\quad = S[1, 2, \cdots, g; 1, 2, \cdots, h \mid i_1, \cdots, i_g, j_1, \cdots, j_h] \\ &\quad - \{S[1, 2, \cdots, g - 1; 1, 2, \cdots, h \mid i_1, \cdots, i_{g-1}, j_1, \cdots, j_h] + \cdots\}. \\ &\quad + \cdots + (-1)^{g+h-1}\{S[1; 0 \mid i_1] + \cdots\}. \end{aligned} \right. \tag{3.9}$$

The estimate for a $(p + q)$th interaction involving the factors (say) $\{A_s\}$, $\{B_r\}$ associated with the respective levels i_s, j_r $(s = 1, 2, \cdots, p;\ r = 1, 2, \cdots, q)$ can conveniently be written as

$$(a_{12\cdots p}\hat{b}_{12\cdots q})_{i_{12}\cdots{}_p j_{12}\cdots{}_q} = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \{w\}, \tag{3.10}$$

where $\{w\}$ denotes the sum of all S-functions involving exactly $w \leqq p + q$ factors from the above set.

3.2. Variances, covariances, and tests of significance. In this section we shall obtain the variances and covariances of the main effects and interaction terms. It will be shown that these can be written as direct products of matrices and this leads directly to the appropriate sums of squares for the analysis of variance. Four lemmas pertaining to the S-functions are derived and are used for proving three basic theorems pertaining to the analysis.

LEMMA 3.1. *The variance of* $S = S[1, 2, \cdots, p; 1, 2, \cdots, q \mid i, j]$ *is*

$$\text{Var } S = \frac{\sigma^2}{r(k-1)v}[(MN-1)(k-c_1) + n(M-1)(c_1-c_2)], \tag{3.11}$$

where $M = \prod_{s=1}^{p} m_s$, $N = \prod_{r=1}^{q} n_r$.

PROOF. The number of treatments summed in S is $v/MN = mn/MN$ which can be regarded as m/M groups of n/N treatments each, such that treatments within the same group are first associates and treatments in different groups are second associates. Then there are $\binom{mn/MN}{2}$ different pairs of treatments among the mn/MN treatments in S, of which

$$\frac{m}{M}\binom{n/N}{2} = \frac{v(n-N)}{2N^2M}$$

are 1st associates and

$$\frac{1}{2}\left(\frac{m}{M} - 1\right)\left(\frac{n}{N}\right)\frac{v}{NM}$$

are 2nd associates. Therefore the variance of S is

$$\left\{\begin{aligned} \text{Var } S = \frac{\sigma^2M^2N^2}{v^2}\Bigg\{&\frac{v}{MN}\left[\frac{vk - [k + c_1(n-1) + c_2n(m-1)}{r(k-1)v}\right] \\ &+ \frac{2v(n-N)}{2N^2M}\left[\frac{c_1v - [k + c_1(n-1) + c_2n(m-1)]}{r(k-1)v}\right] \\ &+ \frac{2v(m-M)n}{2(MN)^2}\left[\frac{c_2v - [k + c_1(n-1) + c_2n(m-1)]}{r(k-1)v}\right]\Bigg\}, \end{aligned}\right. \tag{3.12}$$

which on simplifying gives the desired result.

LEMMA 3.2. *Let* $S = S[1, 2, \cdots, p; 1, 2, \cdots, q \mid i, j]$ *and*

$$S' = S[1', 2', \cdots, p'; 1', 2', \cdots, q' \mid i', j']$$

be two S-functions having a A factors and b B factors in common, such that for a_1 *and* b_1 *of these factors, the levels are identical and for* a_2 *and* b_2 *of these (common) factors, the levels are different* ($a = a_1 + a_2$, $b = b_1 + b_2$). *Then*

$$\text{Cov}(S, S') = \frac{\sigma^2}{r(k-1)v}\left\{\begin{aligned}&(M_1N_1 - 1)(k - c_1) \\ &+ n(M_1 - 1)(c_1 - c_2)\end{aligned}\right\}, \tag{3.13}$$

where $M_1 = \prod_{i=1}^{a_1} m_i$ *(product of the levels of the* a_1 *factors having common levels) and* $N_1 = \prod_{j=1}^{b_1} n_j$ *(product of the levels of the* b_1 *factors having common levels).*

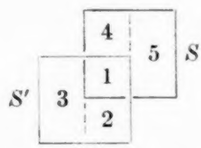
PROOF. The number of treatments summed in S and S' are v/MN and $v/M'N'$ respectively. These treatments can be regarded as consisting of two rectangular treatment arrays of dimensions $(n/N) \times (m/M)$ and

$$(n/N') \times (m/M')$$

respectively. The two arrays will overlap if they have common treatments and the number of such common treatments is

$$\frac{vM_1N_1}{(MN)(M'N')} = \left(\frac{mM_1}{MM'}\right)\left(\frac{nN_1}{NN'}\right).$$

It is convenient to depict the intersection of the rectangular arrays by the five regions as shown below,



where region (1) is an array representing the common treatments having (nN_1/NN') rows and (mM_1/MM') columns. If $\sum (i)$ represent the sum of the treatments in the ith region $(i = 1, 2, 3, 4, 5)$, then

$$\text{(3.14)} \qquad \begin{cases} S = \sum (1) + \sum (4) + \sum (5), \\ S' = \sum (1) + \sum (2) + \sum (3). \end{cases}$$

Hence, in order to find the covariance between S and S', it is necessary to find the number of pairs of 1st and 2nd associates formed from the multiplication of S and S'. These will give pairs formed from $\sum (1)^2$, $\sum (1) \sum (2)$, $\sum (1) \sum (3)$, $\sum (1) \sum (4)$, $\sum (2) \sum (4)$, $\sum (3) \sum (4)$, $\sum (1) \sum (5)$, $\sum (2) \sum (5)$, and $\sum (3) \sum (5)$.

Define

$$\text{(3.15)} \qquad \begin{cases} m_1 = \dfrac{mM_1}{MM'}, & n_1 = \dfrac{nN_1}{NN'}, \\ m_2 = \dfrac{m}{MM'}(M - M_1), & n_2 = \dfrac{n}{NN'}(N - N_1), \\ m_3 = \dfrac{m}{MM'}(M' - M_1), & n_3 = \dfrac{n}{NN'}(N' - N_1). \end{cases}$$

Then the dimensions of the five regions are:

region (1): $n_1 \times m_1$,

region (2): $n_2 \times m_1$,

$$\text{region (3):}\quad (n_1 + n_2) \times m_2,$$

$$\text{region (4):}\quad n_3 \times m_1,$$

$$\text{region (5):}\quad (n_1 + n_3) \times m_3.$$

Since the treatments in the same row are 1st associates of each other and treatments in different rows 2nd associates, it is an easy matter to count the number of 1st and 2nd associates arising from pairs formed from $\sum (i) \sum (j)$. Performing the necessary algebra, we find that the total number of 1st associate pairs is

$$\frac{vM_1(n - N_1)}{(NN')(MM')}$$

and the total number of 2nd associate pairs is

$$\frac{vn(m - M_1)}{(NN')(MM')}.$$

Therefore,

$$(3.16)\quad \left\{\begin{aligned} \operatorname{Cov}(S, S') = \frac{\sigma^2(MM')(NN')}{v^2}&\left\{\frac{vM_1N_1}{(MM'NN')}\ [\operatorname{Var} t]\right.\\ &+ \frac{vM_1(n - N_1)}{(NN'MM')}\ \operatorname{Cov}\ (\text{1st associates})\\ &\left.+ \frac{vn(m - M_1)}{(NN'MM')}\ \operatorname{Cov}\ (\text{2nd associates})\right\}.\end{aligned}\right.$$

On simplifying we get the desired result.

LEMMA 3.3. *Let* $(\hat{ab}) = (a_{12\cdots p}b_{12\cdots q})_{i_{12}\cdots{}_p j_{12}\cdots{}_q}$ *be the estimate of the* $(p + q)$th *factor interaction associated with the factors* $\{A_s\}(s = 1, 2, \cdots, p)$ *and*

$$\{B_r\}(r = 1, 2, \cdots, q).$$

Let $S' = S[1; 2', \cdots, p'; 1', 2', \cdots, q' \mid i', j']$ *be an S-function which is not associated with all factors (regardless of level) of* $(\hat{ab})$. *Then*

$$(3.17)\qquad \operatorname{Cov}[(\hat{ab}), S'] = 0.$$

PROOF. Let α be the number of common A factors between (ab) and S', and α_1 and α_2 $(\alpha = \alpha_1 + \alpha_2)$ be the number of these common factors having the same levels and different levels, respectively. Define β, β_1, and β_2 in the same manner with respect to the B factors. Since the interaction $(\hat{ab})$ can be written in the form

$$(\hat{ab}) = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w\{w\},$$

consider a fixed $\{w\}$ and a particular S-function in $\{w\}$ having the characteristics a, a_1, a_2, b, b_1, b_2 as defined in Lemma 3.2.

Define

$$
(3.18)\quad
\begin{cases}
C(0, 0) = -1, & \\
C(a_1) = \sum \cdots \sum (m_{s_1} m_{s_2} \cdots m_{s_{a_1}} - 1), & a_1 \leqq \alpha_1 , \\
C(b_1) = \sum \cdots \sum (n_{r_1} n_{r_2} \cdots n_{r_{b_1}} - 1), & b_1 \leqq \beta_1 , \\
C(a_1 , b_1) = \sum \cdots \sum (m_{s_1} m_{s_2} \cdots m_{s_{a_1}} n_{r_1} n_{r_2} \cdots n_{r_{b_1}} - 1), & \\
\qquad\qquad a_1 \leqq \alpha_1 , \quad b_1 \leqq \beta_1 , &
\end{cases}
$$

where the summations are only over combinations of A and B factors taken a_1 and b_1 at a time respectively, such that these factors are those in which $(\widehat{ab})$ and S have in common at the same level.

Then the covariance between S' and $\{w\}$ can be written

$$
(3.19)\quad
\begin{cases}
\operatorname{Cov}[S', \{w\}] = \dfrac{\sigma^2}{r(k-1)v} \left\{ \displaystyle\sum_{a_1+b_1=w} \binom{p+q-\alpha-\beta}{w-a_1-b_1} \right. \\
\qquad\qquad \cdot \left[(k-c_1)C(a_1, b_1) + n\dbinom{\beta_1}{b_1}(c_1-c_2)C(a_1) \right] \\
\quad - \displaystyle\sum_{\substack{a_1+b=w \\ b_2 \neq 0}} \binom{p+q-\alpha-\beta}{w-a_1-b} \\
\qquad\qquad \cdot \left[\dbinom{\alpha_1}{a_1}\dbinom{\beta}{b}(k-c_1) - n\dbinom{\beta}{b}(c_1-c_2)C(a_1) \right] \\
\quad \left. + \displaystyle\sum_{a_2=1}^{w} \binom{p+q-\alpha-\beta}{w-a_2}\binom{\alpha_2}{a_2}[(k-c_1)+n(c_1-c_2)] \right\}.
\end{cases}
$$

Note that the first summation is for those S-functions in $\{w\}$ for which

$$a_2 = b_2 = 0;$$

the second summation refers to $a_2 = 0$, $b_2 \neq 0$; and the third summation is when $a_2 \neq 0$. Since the covariance between S' and $(\widehat{ab})$ is

$$
(3.20)\qquad \operatorname{Cov}[S', (\;\hat{}\;)] = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \operatorname{Cov}[S', \{w\}],
$$

we can substitute (3.19) in (3.20) to obtain an explicit expression for (3.20). Now with respect to fixed values of a_1, a_2, b_1, and b_2 the only terms contributing to the first summation in (3.19) is when

$$w = a_1 + b_1, \cdots, p + q + a_1 + b_1 - \alpha - \beta;$$

the value of w contributing to the second summation in (3.19) is for

$$w = a_1 + b, \cdots, p + q + a_1 + b - \alpha - \beta$$

and the contributing value of w for the last summation in (3.19) is when

$$w = a_2, \cdots, p + q + a_2 - \alpha - \beta.$$

Therefore collecting coefficients of

$$\left[(k - c_1)C(a_1, b_1) + n\binom{\beta_1}{b_1}(c_1 - c_2)C(a_1)\right]$$

in (3.20) gives

$$(3.21) \qquad (-1)^{a_1+b_1} \sum_{w=0}^{p+q-\alpha-\beta} \binom{p + q - \alpha - \beta}{w} (-1)^w = 0$$

for all a_1 and b_1. Collecting coefficients of

$$\binom{\beta}{b}\left[\binom{\alpha_1}{a_1}(k - c_1) - n(c_1 - c_2)C(a_1)\right]$$

results in

$$(-1)^{a_1+b+1} \sum_{w=0}^{p+q-\alpha-\beta} \binom{p + q - \alpha - \beta}{w} (-1)^w = 0$$

for all a_1, b_1, and b_2. Finally, with respect to the coefficient of

$$\binom{\alpha_2}{a_2}[(k - c_1) + n(c_1 - c_2)]$$

in (3.20) we have

$$(-1)^{a_1} \sum_{w=0}^{p+q-\alpha-\beta} \binom{p + q - \alpha - \beta}{w} (-1)^w = 0.$$

LEMMA 3.4. *Let* $(\hat{ab}) = (a_{12\cdots p}\hat{b}_{12\cdots q})_{i_{12}\cdots p j_{12}\cdots q}$ *be an estimate of the* $(p + q)$ *factor interaction associated with the factors* $\{A_i\}(i = 1, 2, \cdots, p)$ *and*

$$\{B_j\}(j = 1, 2, \cdots, q).$$

Let $S' = S[1, 2, \cdots, p; 1, 2, \cdots, q \mid i, j]$ *be an S-function associated with the same factors as* $(\hat{ab})$ *such that* α_1 *and* β_1 *of the* A *and* B *factors have common levels. Then,*

$$(3.22) \quad \mathrm{Cov}\,[(\hat{ab}), S'] = \begin{cases} (-1)^{p+q+\alpha_1+\beta_1}\theta(1, 2, \cdots, \alpha_1)\phi(1, 2, \cdots, \beta_1)\dfrac{\sigma^2}{E_b r v}, & \text{if } q \neq 0, \\ (-1)^{p+\alpha_1}\theta(1, 2, \cdots, \alpha_1)\dfrac{\sigma^2}{E_a r v}, & \text{if } q = 0, \end{cases}$$

where

$$(3.23) \quad \begin{aligned} \theta(1, 2, \cdots, \alpha_1) &= (m_1 - 1)(m_2 - 1) \cdots (m_{\alpha_1} - 1), \\ \phi(1, 2, \cdots, \beta_1) &= (n_1 - 1)(n_2 - 1) \cdots (n_{\beta_1} - 1), \end{aligned}$$

and

$$
\left\{
\begin{aligned}
E_a &= \frac{(k-1)}{(k-c_1)+n(c_1-c_2)}, \\
E_b &= \frac{k-1}{k-c_1}.
\end{aligned}
\right. \tag{3.24}
$$

PROOF. If we expand $(a\hat{\ }b)$ in terms of S-functions (Eq. 3.10), we can write the covariance between S' and a fixed $\{w\}$ for $q \neq 0$ as

$$
\begin{aligned}
\operatorname{Cov}[S', \{w\}] = \frac{\sigma^2}{r(k-1)v} \Bigg\{ \Bigg[& (k-c_1) \sum_{a_1+b_1=w} C(a_1, b_1) \\
& + n(c_1-c_2) \sum_{a_1+b_1=w} \binom{\beta_1}{b_1} C(a_1) \Bigg] \\
& - \Bigg[\sum_{\substack{a_2+s=w \\ a_2 \neq 0}} \binom{\alpha_2}{a_2} \binom{q+\alpha_1}{s} [(k-c_1)+n(c_1-c_2)] \Bigg] \\
& - \Bigg[\sum_{\substack{a_1+b_1+b_2=w \\ b_2 \neq 0}} \binom{\alpha_1}{a_1} \binom{\beta_1}{b_1} \binom{\beta_2}{b_2} [(k-c_1)] \\
& - n(c_1-c_2) \sum_{\substack{a_1+b_1+b_2=w \\ b_2 \neq 0}} \binom{\beta_1}{b_1} \binom{\beta_2}{b_2} C(a_1) \Bigg] \Bigg\},
\end{aligned} \tag{3.25}
$$

where the first bracket is when $a_2 = b_2 = 0$; the second bracket is the case $a_2 \neq 0$; and the third bracket refers to $a_2 = 0$, $b_2 \neq 0$. Substituting (3.25) in

$$
\operatorname{Cov}[S', (a\hat{\ }b)] = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \operatorname{Cov}[S', \{w\}] \tag{3.26}
$$

results in the first bracket being written as (neglecting the constant term)

$$
\begin{aligned}
(-1)^{p+q+\alpha_1+\beta_1} \Bigg[& \sum_{w=1}^{\alpha_1+\beta_1} (-1)^w \sum_{a_1+b_1=w} C(a_1, b_1)(k-c_1) \\
& + n(c_1-c_2) \sum_{w=1}^{\alpha_1+\beta_1} (-1)^w \sum_{a_1+b_1=w} \binom{\beta_1}{b_1} C(a_1) \Bigg] \\
= {} & (-1)^{p+q+\alpha_1+\beta_1} \theta(1, 2, \cdots, \alpha_1) \phi(1, 2, \cdots, \beta_1)(k-c_1) \\
& + (-1)^{p+q} n(c_1-c_2) \sum_{a_1=1}^{\alpha_1} C(a_1) \sum_{w=a_1}^{a_1+q} (-1)^w \binom{\beta_1}{w-a_1}.
\end{aligned} \tag{3.27}
$$

With respect to the bracket when $a_2 \neq 0$, we can write these terms after substituting in (3.26) as

$$
[k-c_1+n(c_1-c_2)] \left[\sum_{r=1}^{p+q} (-1)^r \binom{p+q}{r} - \sum_{r=1}^{\alpha_1+q} (-1)^r \binom{\alpha_1+q}{r} \right] = 0. \tag{3.28}
$$

Finally for the terms where $a_2 = 0$, $b_2 \neq 0$, after substituting in (3.26), we can write

$$(3.29)\quad \begin{cases} (-1)^{p+q+1}\left[\sum\limits_{w=1}^{\alpha_1+q}(-1)^w\binom{\alpha_1+q}{w} - \sum\limits_{w=1}^{\alpha_1+\beta_1}(-1)^w\binom{\alpha_1+\beta_1}{w}\right](k-c_1) \\ +\ (-1)^{p+q}n(c_1-c_2)\left[\sum\limits_{a_1=1}^{\alpha_1}C(a_1)\sum\limits_{w=a_1}^{\alpha_1+q}(-1)^w\left[\binom{\beta_1+\beta_2}{w-a_1}-\binom{\beta_1}{w-a_1}\right]\right]. \end{cases}$$

The first term in (3.29) is identically zero and combining the second term of the right hand side of (3.27) with the second term of (3.29) gives

$$(-1)^{p+q}n(c_1-c_2)\left[\sum_{a_1=1}^{\alpha_1}C(a_1)\sum_{r=a_1}^{a_1+q}(-1)^r\binom{q}{r-a_1}\right] = 0.$$

Thus the Lemma is true for $q \neq 0$. For $q = 0$, the covariance between S' and $\{w\}$ will be

$$(3.30)\quad \begin{cases} \operatorname{Cov}[S', \{w\}] = \dfrac{\sigma^2}{r(k-1)v}\Big\{[(k-c_1)+n(c_1-c_2)]C(a_1) \\ \qquad - \sum\limits_{a_1+a_2=w}\dbinom{\alpha_1}{a_1}\dbinom{\alpha_2}{a_2}[(k-c_1)+n(c_1-c_2)]\Big\} \end{cases}$$

and following the same reasoning as for $q \neq 0$, we can prove the Lemma for $q = 0$.

THEOREM 3.1. *Let* $(\hat{ab}) = (a_{12\cdots p}b_{12\cdots q})_{i_{12}\cdots{}_p j_{12}\cdots{}_q}$ *be an estimate of the*

$$(p+q)\text{th}$$

factor interaction associated with the factors $\{A_i\}$ $(i = 1, 2, \cdots, p)$ *and*

$$\{B_j\}\ (j = 1, 2, \cdots, q);$$

let $(\hat{ab})' = (a'_{12\cdots r}b'_{12\cdots s})_{i'_{12}\cdots{}_r j'_{12}\cdots{}_s}$ *be a* $(r + s)$ *factor interaction associated with the factors* $\{A'_i\}$ $(i = 1, 2, \cdots, r)$ *and* $\{B'_j\}$$(j = 1, 2, \cdots, s)$*, such that all factors are not identical between* $(\hat{ab})$ *and* $(\hat{ab})'$*. Then the two different interactions are uncorrelated.*

PROOF. $(\hat{ab})$ can be expanded in terms of S-functions, such that no S-function contains all the factors of $(\hat{ab})'$. Hence by Lemma 3.3, the covariance between all such S-functions and $(\hat{ab})'$ are zero which proves the theorem.

THEOREM 3.2. *The variance of the* $(p + q)$ *factor interaction*

$$(\hat{ab}) = (a_{12\cdots p}b_{12\cdots q})_{i_{12}\cdots{}_p j_{12}\cdots{}_q}$$

associated with the factors $\{A_i\}$ $(i = 1, 2, \cdots, p)$ *and* $\{B_j\}$ $(j = 1, 2, \cdots, q)$ *is*

$$(3.31)\qquad \operatorname{Var}(\hat{ab}) = \begin{cases} \theta(1, 2, \cdots, p)\dfrac{\sigma^2}{E_a r v}, & \text{if } q = 0, \\ \theta(1, 2, \cdots, p)\phi(1, 2, \cdots, q)\dfrac{\sigma^2}{E_b r v}, & \text{if } q \neq 0. \end{cases}$$

PROOF. Using Lemma 3.3, we can show that

$$\operatorname{Var}(\widehat{ab}) = \operatorname{Cov}[(\widehat{ab}), S],$$

where S denotes that S-function coinciding in all factors and levels with the interaction $(\widehat{ab})$. Hence, by Lemma 3.4 the theorem is proved.

THEOREM 3.3. *Let $(\widehat{ab})_{ij}$ and $(\widehat{ab})_{i'j'}$ be two estimates of a $(p + q)$ factor interaction associated with the factors $\{A_i\}$ $(i = 1, 2, \cdots, p)$ and*

$$\{B_j\} \ (j = 1, 2, \cdots, q)$$

such that for α_1 of the A factors and β_1 of the B factors, the levels are identical. Then

$$\operatorname{Cov}[(\widehat{ab})_{ij}, (\widehat{ab})_{i'j'}] \tag{3.32}$$

$$= \begin{cases} (-1)^{p+\alpha_1}\theta(1, 2, \cdots, \alpha_1) \dfrac{\sigma^2}{E_a rv}, & \text{if } q = 0, \\ (-1)^{p+q+\alpha_1+\beta_1}\theta(1, 2, \cdots, \alpha_1)\phi(1, 2, \cdots, \beta_1) \dfrac{\sigma^2}{E_b rv}, & \text{if } q \neq 0. \end{cases}$$

PROOF. Expanding $(\widehat{ab})_{i'j'}$ in terms of S-functions, taking the covariance of $(\widehat{ab})_{ij}$ with each of the S-functions associated with $(\widehat{ab})_{i'j'}$ and using Lemma 3.3, results in

$$\operatorname{Cov}[(\widehat{ab})_{ij}, (\widehat{ab})_{i'j'}] = \operatorname{Cov}[(\widehat{ab})_{ij}, S'],$$

where S' is that S-function associated with the factors $\{A_i\}$ and

$$\{B_j\} \ (i = 1, 2, \cdots, p; j = 1, 2, \cdots, q)$$

and levels $i'_{12\cdots p}j'_{12\cdots q}$.

Hence, by Lemma 3.4 the theorem is proved.

Theorems 3.1 through 3.3 give the variances and covariances of any general interaction. Define the square matrices $M(i)$ and $N(j)$ of dimension m_i and n_j, respectively, by

$$\begin{cases} M(i) = m_i I - J, & i = 1, 2, \cdots, g, \\ N(j) = n_j I - J, & j = 1, 2, \cdots, h, \end{cases} \tag{3.33}$$

where J is a matrix of appropriate dimension having all elements unity. Then the variance-covariance matrix of the estimates of the $(p + q)$ factor interaction $(a_{12\cdots p}\hat{b}_{12\cdots q})_{i_{12\cdots p}j_{12\cdots q}}$ ranging over all the $\prod_{i=1}^{p} m_i \prod_{j=1}^{q} n_j$ combinations is given by the direct matrix product

$$\frac{\sigma^2}{E_a rv}[M(1)*M(2)* \cdots *M(p)], \qquad \text{if } q = 0, \tag{3.34}$$

or

$$\frac{\sigma^2}{E_b rv}[M(1)* \cdots *M(p)*N(1)* \cdots *N(q)], \qquad \text{if } q \neq 0. \tag{3.35}$$

Therefore, since $M(i)^2 = m_i M(i)$, $N(j)^2 = n_j N(j)$, and using Lemmas (2.2) and (2.3) the sums of squares

$$\frac{E_a rv}{\prod_{s=1}^{p} m_s} \sum_{i_{12\cdots p}} (\hat{a}_{12\cdots p})^2_{i_{12\cdots p}},$$

(3.36)

$$\frac{E_b rv}{\prod_{s=1}^{p} m_s \prod_{r=1}^{q} n_r} \sum_{i_{12\cdots p} j_{12\cdots q}} (a_{12\cdots p} \hat{b}_{12\cdots q})^2_{i_{12\cdots p} j_{12\cdots q}}, \qquad \text{if } q \neq 0,$$

follow $\chi^2\sigma^2$ distributions with $\prod_{s=1}^{p} (m_s - 1)$ and $\prod_{s=1}^{p} \prod_{r=1}^{q} (m_s - 1)(n_r - 1)$ degrees of freedom respectively if the null hypothesis of no interaction effect is true.

Using Lemma 2.1 these sums of squares have the expected values

$$\frac{E_a rv}{\prod_{s=1}^{p} m_s} \sum_{i_{12\cdots p}} (a_{12\cdots p})^2_{i_{12\cdots p}} + \prod_{i=1}^{p} (m_i - 1)\sigma^2$$

and

$$\frac{E_b rv}{\prod_{s=1}^{p} m_s \prod_{r=1}^{q} n_r} \sum_{i_{12\cdots p} j_{12\cdots q}} (a_{12\cdots p} b_{12\cdots q})^2 + \prod_{i=1}^{p} (m_i - 1) \prod_{j=1}^{q} (n_j - 1)\sigma^2.$$

The entire intra-block analysis of variance can be summarized in Table 1 where B represents the $b \times 1$ vector of block totals, Q is the $v \times 1$ vector of adjusted treatment totals, $\hat{t}$ is the $v \times 1$ vector of treatment estimates,

$$g = \frac{(\text{grand total})^2}{bk},$$

and terms such as

$$\sum_{i_{12\cdots p}} (\hat{a}_{12\cdots p})^2_{i_{12\cdots p}}$$

are written as $(\hat{a}_{12\cdots p})^2$, etc.

The computations for the analysis of variance are straightforward and actually amount to treating the $\hat{t}_u$'s as observations on a one replicate factorial experiment, where all sums of squares are multiplied by $E_a r$ or $E_b r$. It is also clear from the analysis of variance that the various interactions are estimated with one of a possible two types of efficiencies. If the interaction is composed only of A factors the efficiency is E_a, otherwise the efficiency will be E_b.

Extension to the balanced incomplete block designs. The balanced incomplete block designs can also be used for asymmetric factorial arrangements by assigning the v factorial combinations to the v treatments of the balanced incomplete block design. All results immediately follow by regarding the balanced incomplete block designs as a "degenerate" partially balanced design. Then

TABLE 1

Summary of intra-block analysis of variance

Source	Degrees of freedom	Sum of squares
Blocks (unadjusted)	$b - 1$	$\dfrac{B'B}{k} - g$
Treatments (adjusted)	$v - 1$	$\hat{t}'Q$
A_1	$(m_1 - 1)$	$\dfrac{v}{m_1} E_a r(\hat{a}_1)^2$
A_2	$(m_2 - 1)$	$\dfrac{v}{m_2} E_a r(\hat{a}_2)^2$
$\vdots$	$\vdots$	$\vdots$
B_1	$(n_1 - 1)$	$\dfrac{v}{n_1} E_b r(\hat{b}_1)^2$
$\vdots$	$\vdots$	$\vdots$
A_1B_1	$(m_1 - 1)(n_1 - 1)$	$\dfrac{v}{m_1 n_1} E_b r(\widehat{a_1 b_1})^2$
$\vdots$	$\vdots$	$\vdots$
$A_1A_2 \cdots A_g B_1 \cdots B_h$	$\prod_{s=1}^{g} (m_s - 1) \prod_{r=1}^{h} (n_r - 1)$	$\dfrac{v}{\prod_1^g m_s \prod_1^h n_r} E_b r(\widehat{a_{12\cdots g} b_{12\cdots h}})^2$
Error	$n_e = bk - b - v + 1$	$S_e = \sum_{i,j} y_{ij}^2 - t'Q - \dfrac{B'B}{k}$
Total	$bk - 1$	$\sum_{i,j} y_{ij}^2 - g$

$c_1 = c_2 = k/v$ in (3.4), $E_a = E_b = E = v(k-1)/k(v-1)$, and all main effects and interactions are estimated with an efficiency factor E.

4. The recovery of inter-block information. If the block effects b_j in (3.4) can be regarded as a sequence of random variables such that

$$
\begin{cases} E(b_j) = 0, \qquad \operatorname{Var} b_j = \sigma_b^2, & \\ \operatorname{Cov}(b_j, b_{j'}) = 0, & \text{for } j \neq j', \\ \operatorname{Cov}(\epsilon_{ij}, b_{j'}) = 0, & \end{cases} \tag{4.1}
$$

it will be possible to extract additional information from the block totals. This additional analysis is sometimes termed the recovery of inter-block information or the interblock analysis. With respect to the balanced incomplete block designs, the inter-block analysis was first developed by Yates [20] and appears in the books by Cochran and Cox [9], Federer [10], and Kempthorne [13]. The inter-block analysis with respect to the partially balanced designs is discussed in a particularly simple form by Bose and Shimamoto [5] and by Bose, Clatworthy, and Shrikhande [2]. Generally it will be possible to use this inter-block information in two ways. The preceding references discuss how one may combine the inter-block information with the intra-block information in order to obtain the most precise treatment estimates. The paper by Zelen [21] discusses how one can use this inter-block information for obtaining additional independent tests of significance for every hypothesis pertaining to the treatments.

Define $Q_i^* = T_i - Q_i - r\bar{y}$, where Q_i is the ith adjusted treatment total, T_i is the total for the ith treatment and $\bar{y}$ is the grand average of all measurements. Then the best estimates of the treatments using both the intra- and inter-block information can be written as

$$\bar{t}_i = \frac{1}{R(k-1)} \{kP_i + d_1 S_1(P_i) + d_2 S_2(P_i)\}, \tag{4.2}$$

where

$$\begin{cases} P_i = WQ_i + W^*Q_i^*, \\ R = r\left[W + \dfrac{W^*}{k-1}\right], \\ W = \dfrac{1}{\sigma^2}, \qquad W^* = \dfrac{1}{\sigma^2 + k\sigma_b^2}. \end{cases} \tag{4.3}$$

The constants d_1, d_2 are usually tabulated with all the designs. Note that (4.2) is the same form as (3.4) except that P_i replaces Q_i, R replaces r, and d_1, d_2 replace c_1, c_2. Thus all results in Section 3 carry over directly by substituting the above changes in the formulas of Section 3 and replacing σ^2 by unity.

On the other hand, under certain conditions which are elaborated in [4], one can also obtain additional independent tests of significance using only the inter-block information. Three cases have to be considered depending on whether the group divisible design is a regular, singular, or semi-regular design. These are the three exhaustive classes of group divisible designs introduced by Bose and Connor [3].

For the regular group divisible designs the inter-block treatment estimates can be written as

$$t_i^* = \frac{kQ_i^* + c_1^* S_1(Q_i^*) + c_2^* S_2(Q_i^*)}{r} \tag{4.4}$$

and will have a variance of

$$\text{Var } t_i^* = \left[\frac{vk - [k + (n-1)c_1^* + n(m-1)c_2^*]}{rv}\right](\sigma^2 + k\sigma_b^2).$$

Also if t_i^* and t_j^* are sth associates $(s = 1, 2)$,

$$\text{Cov }(t_i^*, t_j^*) = \left[\frac{vc_s^* - [k + (n-1)c_1^* + n(m-1)c_2^*]}{rv}\right](\sigma^2 + k\sigma_b^2).$$

The quantities c_1^* and c_2^* are defined by

$$c_s^* = \frac{c_s\Delta - r\lambda_s}{\Delta - rH + r^2} \qquad (s = 1, 2),$$

where the parameters c_s, Δ, λ_s, and H are the usual parameters for partially balanced designs (cf. [2], [5]). Therefore all results for Section 3 apply directly by replacing σ^2 by $(\sigma^2 + k\sigma_b^2)$, c_s by c_s^*, and $r(k-1)$ by r. This results in the two efficiencies being

$$E_a^* = \frac{1}{k - c_1^* + n(c_1^* - c_2^*)},$$

$$E_b^* = \frac{1}{k - c_1^*},$$

and the breakdown of the $v-1$ treatment sum of squares, using only the inter-block information, is similar to Table 1. If $b > v$, there will be an independent estimate of $\sigma^2 + k\sigma_b^2$, thus permitting independent inter-block tests of significance for the main effects and interactions.

With respect to the singular designs, the intra-block efficiencies are

$$E_a < 1, \qquad E_b = 1.$$

Hence it is only possible to obtain inter-block estimates for those main effects and interactions associated only with the A factors. Since treatments in the same group are first associates, $1/n[t_i + S_1(t_i)]$ represents the average of the group to which treatment i belongs. This average is estimated by

$$\frac{1}{n}[t_i^* + S_1(t_i^*)] = \frac{{}^{z}Q_i^*}{E_a^* r}, \qquad E_a^* = \frac{mn - k}{k(m-1)}. \tag{4.5}$$

There will be m such estimates, thus making it possible to have S-functions of the form $S[1, 2, \cdots, p; 0 \mid i]$ for $p \leqq g$. Then all results of Section 3 follow by replacing E_a by E_a^* and σ^2 by $\sigma^2 + k\sigma_b^2$. If $b > m$, this will permit an estimate of $\sigma^2 + k\sigma_b^2$ and thus we can have independent inter-block tests of significance for the A factor.

The semi-regular group divisible designs have the intra-block efficiencies $E_a = 1$, $E_b < 1$. Therefore it is possible to obtain inter-block estimates of those main effects and interactions having the intra-block efficiency E_b. These $(v - m)$ contrasts can be found by using the normal equations

$$\sum_{s=1}^{v} \frac{\lambda_{is}}{k} t_s^* = Q_i^*, \qquad i = 1, 2, \cdots, v, \tag{4.6}$$

where $\lambda_{ii} = r$, and λ_{is} = number of blocks in which treatments i and s appear together. The rank of Eqs. (4.6) is exactly $(v - m)$. If $b > (v - m)$, then it will be possible to have an independent estimate of $(\sigma^2 + k\sigma_b^2)$, thus allowing independent inter-block tests of significance for testing these contrasts. An open problem is to simplify this analysis.

Extension to balanced incomplete block designs. Similar results apply to the recovery of inter-block information for the balanced incomplete block designs. The best combined estimate can be written as

$$\bar{t}_i = \frac{P_i}{r[WE + W^*(1-E)]} = \frac{P_i}{\bar{E}R}, \qquad \bar{E} = \frac{R(k-1) + \lambda(W - W^*)}{R}.$$

Therefore all results of Section 3 can also be carried over by substituting unity for σ^2, R for r, etc. This produces an efficiency of

$$\bar{E} = \frac{R(k-1) + \lambda(W_{\cdot} - W^*)}{R}.$$

In addition if one wished to obtain additional independent significance tests using the inter-block information only, the treatment estimates can be written

$$t_i^* = \frac{Q_i^*}{(1-E)r}$$

and all results of Section 3 follow by replacing $\sigma^2 + k\sigma_b^2$ for σ^2, and

$$E_a = E_b = 1 - E.$$

Again we will have two independent tests of significance for testing every null hypothesis pertaining to the main effects and interactions.

Acknowledgement. The author is indebted to Professor R. C. Bose and Dr. Max Halperin for drawing the author's attention to Lemmas 2.1 and 2.2, respectively. Also thanks go to Dr. Alan Goldman for reading a preliminary version of this paper.

REFERENCES

[1] R. C. Bose, "Mathematical theory of the symmetric factorial design," *Sankhyā*, Vol. 8 (1947), pp. 107–166.

[2] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, *Tables of Partially balanced designs with two associate classes*, Institute of Statistics of Univ. of North Carolina, Reprint series No. 50 (1954).

[3] R. S. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[4] R. C. Bose and K. Kishen, "On the problem of confounding in general symmetrical factorial designs," *Sankhyā*, Vol. 5 (1940), pp. 21–26.

[5] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[6] R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 167–195.

[7] W. H. Clatworthy, "Partially balanced incomplete block designs with two associate classes and two treatments per block," *Journal of Research of the National Bureau of Standards*, Vol. 54 (1955), pp. 177–190.

[8] W. H. Clatworthy, *Contributions on partially balanced incomplete block designs with two associate classes*, National Bureau of Standards Applied Mathematics Series 47, 1956.

[9] W. G. Cochran and G. M. Cox, *Experimental Designs*, New York, John Wiley and Sons, 1950.

[10] W. T. Federer, *Experimental Designs*, New York, Macmillan, 1955.

[11] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh.

[12] R. A. Fisher, "The theory of confounding in factorial experiments in relation to the theory of groups," *Ann. Eugenics*, Vol. 11 (1942), pp. 341–353.

[13] O. Kempthorne, *The Design and Analysis of Experiments*, New York, John Wiley and Sons, 1952.

[14] C. Y. Kramer and R. A. Bradley, "Intra-block analysis for factorials in two-associate class group divisible designs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 349–361.

[15] J. Li, "Design and statistical analysis of some confounded factorial experiments," *Iowa Agricultural Experimental Station Research Bulletin*, *333*, 1934.

[16] K. R. Nair and C. R. Rao, "Confounding in asymmetrical factorial experiments," *J. Roy. Stat. Soc.*, Series B, Vol. 10 (1948), pp. 109–131.

[17] C. R. Rao, "General methods of analysis for incomplete block designs," *J. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 541–561.

[18] C. R. Rao, "On the recovery of inter-block information in varietal trials," *Sankhyā*, Vol. 17 (1956), pp. 105–114.

[19] F. Yates, *The design and analysis of factorial experiments*, Imperial Bureau Soil Science, Harpenden, England, 1937.

[20] F. Yates, "The recovery of inter-block information in balanced incomplete block designs," *Ann. Eugenics*, Vol. 10 (1940), pp. 317–325.

[21] M. Zelen, "The analysis of incomplete block designs," *J. Amer. Stat. Assn.*, Vol. 52 (1957), pp. 204–217.

[22] M. Zelen, "The use of incomplete block designs for factorial experiments" (abstract), *J. Amer. Stat. Assn.*, Vol. 50 (1955), p. 599.

ACCELERATED STOCHASTIC APPROXIMATION[1]

By Harry Kesten

Cornell University

1. Summary. Using a stochastic approximation procedure $\{X_n\}$, $n = 1, 2, \cdots$, for a value θ, it seems likely that frequent fluctuations in the sign of $(X_n - \theta) - (X_{n-1} - \theta) = X_n - X_{n-1}$ indicate that $|X_n - \theta|$ is small, whereas few fluctuations in the sign of $X_n - X_{n-1}$ indicate that X_n is still far away from θ. In view of this, certain approximation procedures are considered, for which the magnitude of the nth step (i.e., $X_{n+1} - X_n$) depends on the number of changes in sign in $(X_i - X_{i-1})$ for $i = 2, \cdots, n$. In theorems 2 and 3,

$$X_{n+1} - X_n$$

is of the form $b_n Z_n$, where Z_n is a random variable whose conditional expectation, given $X_1, \cdots, X_n$, has the opposite sign of $X_n - \theta$ and b_n is a positive real number. b_n depends in our processes on the changes in sign of

$$X_i - X_{i-1}(i \leqq n)$$

in such a way that more changes in sign give a smaller b_n. Thus the smaller the number of changes in sign before the nth step, the larger we make the correction on X_n at the nth step. These procedures may accelerate the convergence of X_n to θ, when compared to the usual procedures ([3] and [5]). The result that the considered procedures converge with probability one may be useful for finding optimal procedures. Application to the Robbins-Monro procedure (Theorem 2) seems more interesting than application to the Kiefer-Wolfowitz procedure (Theorem 3).

2. Statement of the theorem. The formulation of the theorem is similar to that of the theorem given by Dvoretzky [2]. Let θ be a real number and

$$T_n(n = 1, 2, \cdots)$$

be measurable transformations. Let X_1 and $Y_n(n = 1, \cdots)$ be random variables[2] and $\{a_n\}$ a sequence of positive numbers and define

$$X_{n+1}(\omega) = T_n(X_1(\omega), \cdots, X_n(\omega)) + b_n(\omega)Y_n(\omega). \tag{1}$$

The sequence $\{b_n(\omega)\}$ is selected in the following way from the sequence $\{a_n\}$

Received March 28, 1957.

[1] Note added in proof: The author learned recently that investigation of the above procedure had been suggested by Professor H. Robbins long ago.

[2] X_n, Y_n, and Z_n denote random variables, whereas x_n is used to denote values taken by the random variables.

$$\begin{aligned} b_1 &= a_1, \\ b_2 &= a_2, \\ b_n &= a_{t(n)}, \end{aligned} \tag{2}$$

where

$$t(n) = 2 + \sum_{i=3}^{n} \iota[(X_i - X_{i-1})(X_{i-1} - X_{i-2})] \tag{3}$$

and

$$\iota(x) = \begin{cases} 1 & \text{if} \quad x \leqq 0 \\ 0 & \text{if} \quad x > 0. \end{cases}$$

Thus, every time $(X_i - X_{i-1})$ differs in sign from $(X_{i-1} - X_{i-2})$ we take another a_n.

Let $\alpha_n(x_1, \cdots, x_n)$, $\beta_n(x_1, \cdots, x_n)$, $\gamma_n(x_1, \cdots, x_n)$ be nonnegative functions and put

$$\epsilon_N = \sup_{\{x_k\}} \sum_{n=N}^{\infty} \beta_n(x_1, \cdots, x_n), \tag{4}$$

$$\rho(\delta) = \inf_{n=1,2,\cdots} \inf_{\substack{|x_n - \theta| \geqq \delta \\ x_1, \cdots, x_{n-1} \text{ arbitrary}}} \frac{\gamma_n(x_1, \cdots, x_n)}{b_n}. \tag{5}$$[3]

THEOREM 1. *If*

$$| T_n(x_1, \cdots, x_n) - \theta | \leqq \begin{cases} (1 + \beta_n(x_1, \cdots, x_n)) | x_n - \theta | \\ - \gamma_n(x_1, \cdots, x_n) \textit{ when } (T_n - \theta)(x_n - \theta) > 0 \\ \alpha_n(x_1, \cdots, x_n) \textit{ when } (T_n - \theta)(x_n - \theta) \leqq 0, \end{cases} \tag{6}$$

(7) $\lim_{t(n) \to \infty} \alpha_n(x_1, \cdots, x_n) = 0$ *uniformly, for all sequences* $x_1, x_2, \cdots$ *with* $t(n) \to \infty$,

(8)[3] $\lim_{n \to \infty} \frac{(x_n - \theta)\beta_n(x_1, \cdots, x_n)}{b_n} = 0$ *uniformly, for all sequences* $x_1, x_2, \cdots$,

and

$$\lim_{N \to \infty} \epsilon_N = 0, \tag{9}$$

$$\rho(\delta) > 0 \quad \textit{for every positive } \delta, \tag{10}$$

$$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \qquad \textit{and} \qquad a_{n+1} \leqq a_n, \tag{11}$$

[3] In (5), (8), and in (13), b_n depends on $x_1, \cdots, x_n$ as given in (2).

(12)[4] $E(Y_n \mid X_1, \cdots, X_n) = 0,$

$$E(Y_n^2 \mid X_1, \cdots, X_n) \leqq \sigma^2 \textit{ with probability } 1,$$

(13)[3]
$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{\substack{0<|x_n-\theta|\leqq\tau \\ x_1,\cdots,x_{n-1} \text{ arbitrary}}} \cdot P\{T_n(X_1, \cdots, X_n) + b_n Y_n \geqq X_n \mid X_1 = x_1, \cdots, X_n = x_n\} > 0,$$

$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{\substack{0<|x_n-\theta|\leqq\tau \\ x_1,\cdots,x_{n-1} \text{ arbitrary}}} \cdot P\{T_n(X_1, \cdots, X_n) + b_n Y_n < X_n \mid X_1 = x_1, \cdots, X_n = x_n\} > 0,$$

then X_n converges to θ with probability 1.

PROOF OF CONVERGENCE. Without loss of generality we take $\theta = 0$. Also we assume in the following $E \mid X_1 \mid < \infty$. This can be done, because replacing X_1 by

$$X_1^1 = \begin{cases} X_1 & \text{if } \mid X_1 \mid < A \\ A & \text{if } \mid X_1 \mid \geqq A \end{cases}$$

changes the process only with a probability equal to

$$P\{\mid X_1 \mid > A\}.$$

By taking A large enough, this probability becomes arbitrary small. We frequently do not write all the arguments of the functions, e.g., we write β_n for $\beta_n(x_1, \cdots, x_n)$. We shall first prove several lemmas. From

$$E(Y_n \mid X_1, \cdots, X_n) = 0$$

and $E(Y_n^2 \mid X_1, \cdots, X_n) \leqq \sigma^2$ follows immediately.

LEMMA 1. *There exists a function $p(\delta)$ with $0 < p(\delta) < 1$ for $\delta > 0$, and such that*

$$P\left\{Y_n \geqq \frac{\delta}{2} > 0 \mid X_1, \cdots, X_n\right\} \leqq 1 - p(\delta) < 1,$$

$$P\left\{Y_n \leqq -\frac{\delta}{2} < 0 \mid X_1, \cdots, X_n\right\} \leqq 1 - p(\delta) < 1.$$

LEMMA 2.

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \geqq \frac{-\rho^1(\delta) b_n}{2} \middle| X_1, \cdots, X_n; \quad X_n \geqq \delta \text{ and } t(n) \geqq k\right\} \leqq 1 - p(\rho^1(\delta)),$$

[4] $P\{\cdot|\cdot\}$ and $E(\cdot|\cdot)$ denote conditional probabilities and conditional expectations respectively.

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \leqq \frac{\rho^1(\delta)b_n}{2} \middle| X_1, \cdots, X_n; \quad X_n \leqq -\delta \text{ and } t(n) \geqq k\right\} \leqq 1 - p(\rho^1(\delta)),$$

where

$$\rho^1(\delta) = \begin{cases} \rho(\delta) & \text{when } \rho(\delta)a_k \leqq \delta \\ \dfrac{\delta}{a_k} & \text{when } \rho(\delta)a_k > \delta. \end{cases}$$

PROOF. Since $t(n) \geqq k$, we have $b_n \leqq a_k$ and for $X_n \geqq \delta$

$$X_{n+1} \leqq \max[0, (1 + \beta_n)X_n - \gamma_n] + b_n Y_n \leqq X_n + b_n\left[\frac{\beta_n X_n}{b_n} - \rho^1(\delta)\right] + b_n Y_n = X_n + b_n\left[\frac{\beta_n}{b_n} X_n - \rho^1(\delta) + Y_n\right].$$

So by (8) and Lemma 1, we have

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \geqq -\frac{\rho^1(\delta)b_n}{2} \middle| X_1, \cdots, X_n; \quad X_n \geqq \delta, \quad t(n) \geqq k\right\}$$
$$\leqq \liminf_{n\to\infty} P\left\{Y_n \geqq \frac{\rho^1(\delta)}{2} - \epsilon \middle| X_1, \cdots, X_n; \quad X_n \geqq \delta, \quad t(n) \geqq k\right\}$$

for every $\epsilon > 0$.

Application of Lemma 1 gives the first inequality. Similarly we prove the second part of the lemma.

LEMMA 3. *For every k and N*

$$P\{t(n) = k \quad \text{for} \quad n \geqq N \quad \text{and} \quad X_n \nrightarrow 0\} = 0$$

(i.e., *when $X_{n+1} - X_n$ does only change sign a finite number of times, then X_n converges to 0*).

PROOF. When $t(n)$ is constant for $n \geqq N$, then X_n is monotonic for $n \geqq N$. Therefore $\{X_n\}$ converges (possibly to $+\infty$ or $-\infty$). Let the limit be positive, say X. But by Lemma 2 for every $\delta > 0$ and $\epsilon < [\rho^1(\delta)a_k]/2$,

$$\lim_{N\to\infty} P\{X_{n+1} - X_n > -\epsilon \quad \text{and} \quad X_n \geqq \delta \quad \text{and} \quad t(n) \geqq k \quad \text{for all } n \geqq N \cdot | \delta \leqq X_N \quad \text{and} \quad t(N) \geqq k\} = 0,$$

so the probability that $X > \delta$ is zero. Similarly the probability $X < -\delta$ is zero. Since δ is an arbitrary positive number, this proves the lemma.

This lemma allows us to limit ourselves in the sequel to those sequences with $t(n) \to \infty$ and therefore $b_n \to 0$.

LEMMA 4. *Let δ be a fixed positive number. Then there exist positive numbers*

n_0 and t_0 such that, whenever $n \geqq n_0$, $t(n) \geqq t_0$ and $|X_n| \geqq \delta$, one has

$$E\{|X_{n+1}| \mid X_1, \cdots, X_n; \quad |X_n| \geqq \delta\} \leqq |X_n| - \frac{b_n}{4}\rho(\delta).$$

PROOF. Choose t_0 such that $\alpha_n(X_1, \cdots, X_n) \leqq \delta/2$ for $t(n) \geqq t_0$ and

$$a_{t_0} \leqq \min\left(\frac{2\delta}{4\sigma + \rho(\delta)}, \frac{\delta\rho(\delta)}{16\sigma^2}, \frac{\delta}{\rho(\delta)}\right). \tag{14}$$

Then $b_n \leqq a_{t_0}$ for $t(n) \geqq t_0$. We distinguish two cases

$$|T_n(X_1, \cdots, X_n)| \leqq \frac{\delta}{2}. \tag{a}$$

$$E\left\{|X_{n+1}| \mid X_1, \cdots, X_n; |X_n| \geqq \delta, |T_n| \leqq \frac{\delta}{2}\right\}$$

$$\leqq \frac{\delta}{2} + b_n E|Y_n| \leqq \frac{\delta}{2} + b_n\sigma \leqq \delta - \frac{b_n\rho(\delta)}{4} \leqq |X_n| - \frac{b_n\rho(\delta)}{4},$$

$$|T_n(X_1, \cdots, X_n)| > \frac{\delta}{2}. \tag{b}$$

As $\alpha_n(X_1, \cdots, X_n) \leqq \delta/2$ for $t(n) \geqq t_0$, we must have $T_n \cdot X_n > 0$ (cf. (6)). Let $X_n \geqq \delta$. Denote the distribution function of $Y_n(X_1, \cdots, X_n)$ by $H_n(y \mid X)$. As $X_{n+1} = T_n + b_n Y_n$, we have by (12) and (14)

$$E\left\{|X_{n+1}| \,\middle|\, X_1, \cdots, X_n; X_n \geqq \delta, T_n \geqq \frac{\delta}{2}\right\}$$

$$= \int_{-T_n/b_n}^{\infty} (T_n + b_n y)\, dH_n(y \mid X) - \int_{-\infty}^{-T_n/b_n} (T_n + b_n y)\, dH_n(y \mid X)$$

$$\leqq T_n + b_n\left[\int_{-T_n/b_n}^{\infty} y\, dH_n(y \mid X) - \int_{-\infty}^{-T_n/b_n} y\, dH_n(y \mid X)\right]$$

$$= T_n - 2b_n \int_{-\infty}^{-T_n/b_n} y\, dH_n(y \mid X)$$

$$\leqq T_n + 2b_n\left[\int_{-\infty}^{-T_n/b_n} y^2\, dH_n(y \mid X) \int_{-\infty}^{-T_n/b_n} dH_n(y \mid X)\right]^{1/2}$$

$$\leqq T_n + 2b_n\sigma\frac{b_n\sigma}{T_n} \leqq T_n + \frac{4b_n^2\sigma^2}{\delta} \leqq T_n + b_n\frac{\rho(\delta)}{4}.$$

But by (8),

$$|T_n| \leqq |X_n| + b_n\left\{\frac{\beta_n|X_n|}{b_n} - \rho(\delta)\right\} \leqq |X_n| - b_n\frac{\rho(\delta)}{2}$$

for sufficiently large n, say $n \geqq n_0$. For $X_n \leqq -\delta$, the proof is similar. Thus, in all cases

$$E\left\{|X_{n+1}|\,\middle|\,X_1, \cdots, X_n; |X_n| \geqq \delta\right\} \leqq |X_n| - \frac{b_n\rho(\delta)}{4}.$$

LEMMA 5. *For every* $0 < \delta < \delta' < \delta''$

$$P\{\delta < \liminf |X_n| < \delta' \quad \textit{and} \quad \delta'' < \limsup |X_n| \quad \textit{and} \quad t(n) \to \infty\} = 0.$$

PROOF. Choose t_0 and n_0, corresponding to δ as introduced in the preceding lemma. Assume now

$$P\{\delta < \liminf |X_n| < \delta' \quad \text{and} \quad \delta'' < \limsup |X_n| \quad \text{and} \quad t(n) \to \infty\} > 0.$$

Then there exist an $n_1 \geqq n_0$ and $t_1 \geqq t_0$ such that

$$P\{\delta < \liminf |X_n| < \delta' \quad \text{and} \quad \delta'' < \limsup |X_n| \quad \text{and} \quad |X_n| > \delta \quad \text{for all} \quad n \geqq n_1 \quad \text{and} \quad b_{n_1} = a_{t_1}\} > 0. \tag{15}$$

Now introduce a new process.

$$Z_i = |X_i| \quad \text{if} \quad i = 1, \cdots, n_1$$

and

$$Z_{n_1+i} = \begin{cases} |X_{n_1+i}| & \text{if} \quad \delta < Z_{n_1+j} \quad \text{for} \quad j = 0, 1, \cdots, i-1 \quad \text{and} \quad b_{n_1} = a_{t_1}, \\ 0 & \text{otherwise.} \end{cases}$$

Unless $b_{n_1} \neq a_{t_1}$ or $|X_j| \leqq \delta$ for some $j \geqq n_1$, we have always $Z_i = |X_i|$, and thus, by (15), also

$$P\{\delta \leqq \liminf Z_n < \delta' \quad \text{and} \quad \delta'' < \limsup Z_n\} > 0. \tag{16}$$

But by Lemma 4

$$0 \leqq E(Z_{n+1} | X_1, \cdots, X_n) \leqq EZ_n \qquad \text{for} \quad n \geqq n_1.$$

So application of the semimartingale convergence theorem (Loève [4], p. 393) shows that (16) cannot be true. This proves the lemma.

LEMMA 6. $P\{\liminf |X_n| = \infty$ *and* $t(n) \to \infty\} = 0$.

PROOF. If the proposition were not true, we could find, analogous to the last proof, a process Z_i with

$$0 \leqq E(Z_{n+1} | X_1, \cdots, X_n) \leqq EZ_n \quad \text{for sufficiently large } n, \quad \text{say} \quad n \geqq n_2, \tag{17}$$

and

$$P\{\liminf Z_n = \infty\} > 0. \tag{18}$$

But as we took $E|X_1| < \infty$, one would have

$$E|Z_{n_2}| < \infty. \tag{19}$$

However, (17) and (19) together are in contradiction with (18). This proves the lemma.

From Lemmas 3, 5, and 6 one may conclude that with probability 1 either that $\liminf |X_n| = 0$ or $|X_n|$ converges to a finite positive number. We now prove that the last possibility has probability zero.

LEMMA 7.

$$P\{|X_n| \text{ converges to } X \text{ and } 0 < \delta < X < \delta' < \infty \text{ and } t(n) \to \infty\} = 0.$$

PROOF. Choose n_0 and t_0 corresponding to δ, as introduced in Lemma 4.

Assume now

$$P\{|X_n| \text{ converges to } X \text{ and } 0 < \delta < X < \delta' < \infty \text{ and } t(n) \to \infty\} > 0. \tag{20}$$

Again there exist an $n_1 \geqq n_0$ and a $t_1 \geqq t_0$ such that

$$P\{\delta < |X_n| < \delta' \text{ for all } n \geqq n_1 \text{ and } b_{n_1} = a_{t_1}\} > 0 \text{ and } a_{t_1}\rho(\delta) \leqq \delta. \tag{21}$$

By Lemma 2 we can choose n_1 and t_1 so that at the same time for $n \geqq n_1$

$$P\left\{X_{n+1} - X_n \geqq -\frac{\rho(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n;\ X_n \geqq \delta, t(n) \geqq t_1\right\} \leqq 1 - \frac{p(\rho(\delta))}{2}, \tag{22}$$

$$P\left\{X_{n+1} - X_n \leqq \frac{\rho(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n;\ X_n \leqq -\delta, t(n) \geqq t_1\right\} \leqq 1 - \frac{p(\rho(\delta))}{2}. \tag{23}$$

As before we construct a new process.

$$Z_i = |X_i| \quad \text{if} \quad i = 1, \cdots, n_1$$

and

$$Z_{n_1+i} = \begin{cases} |X_{n_1+i}| \text{ if } \delta < Z_{n_1+j} < \delta' & \text{for } j = 0, \cdots, i-1 \text{ and } b_{n_1} = a_{t_1}, \\ Z_{n_1+i-1} - a_{t_1+i-1}\dfrac{\rho(\delta)}{4} & \text{otherwise.} \end{cases}$$

From (21) follows

$$P\{\delta < Z_n < \delta' \text{ for all } n \geqq n_1\} > 0, \tag{24}$$

and thus,

$$P\left\{\left|\sum_{k=n_1}^{n} (Z_{k+1} - Z_k)\right| < 2(\delta' - \delta) \text{ for all } n \geqq n_1\right\} > 0. \tag{25}$$

Denote

$$E(Z_{k+1} - Z_k \mid X_1, \cdots, X_k) \quad \text{by} \quad m_k(X_1, \cdots, X_k) \qquad (= m_k \text{ for short}).$$

By Lemma 4 and the construction of the Z-process,

$$m_k(X_1, \cdots, X_k) \leqq -\frac{c_k}{4}\rho(\delta) \text{ for } k \geqq n_1, \tag{26}$$

where

$$c_{n_1+i} = \begin{cases} b_{n_1+i} & \text{if } \delta < Z_j < \delta' \quad \text{for} \quad j = 0, \cdots, i \quad \text{and} \quad b_n = a_{t_1}, \\ a_{n_1+i} & \text{otherwise.} \end{cases}$$

Further for $k \geqq n_1$,

$$\operatorname{var}(Z_{k+1} - Z_k \mid X_1, \cdots, X_k) \leqq c_k^2\left[\frac{\rho^2(\delta)}{16} + EY_k^2\right] \leqq c_k^2 C, \tag{27}$$

where

$$C = \frac{\rho^2(\delta)}{16} + \sigma^2.$$

In addition

$$\sum_{k=n_1}^{n} c_k \geqq \sum_{k=0}^{n-n_1} a_{t_1+k}. \tag{28}$$

By (25),

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1} - Z_k - m_k)\right| \geqq \left|\sum_{k=n_1}^{n} m_k\right| - 2(\delta' - \delta) \qquad \text{for all } n \geqq n_1\right\} > 0, \tag{29}$$

and thus, by (26) and (28),

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1} - Z_k - m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta) \qquad \text{for all } n \geqq n_1\right\} > 0. \tag{30}$$

But for

$$\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta) > 0$$

we have by Tchebycheff's inequality and (27)

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1}-Z_k-m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k}-2(\delta'-\delta)\right\}$$

(32)
$$\leqq \frac{C\sum_{k=n_1}^{n} Ec_k^2}{\left\{\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k}-2(\delta'-\delta)\right\}^2},$$

(32)
$$\sum_{k=n_1}^{n} Ec_k^2 \leqq \sum_{k=0}^{n-n_1} a_{t_1+k}^2 Er_{t_1+k},$$

where

$$r_{t_1+k} = \text{number of times } c_{n_1+i} = a_{t_1+k}.$$

As soon as the Z_i process differs from the $|X_i|$ process, we don't keep the same a_{t_1+k} for more than one step. Therefore $Er_{t_1+k} \leqq 1 +$ expected number of times that $\{c_{n_1+i} = a_{t_1+k}$ and $\delta < Z_{n_1+j} < \delta'$ for $j = 1, \cdots, i$ and $b_{n_1} = a_{t_1}\}$ occurs.

If $\delta < X_{n_1+i} < \delta'$, then by (22),

$$P\{X_{n_1+i+1} > X_{n_1+i} \mid \delta < |X_{n_1+j}| < \delta' \quad j = 1, \cdots, i$$
$$\text{and } b_{n_1} = a_{t_1} \text{ and } X_{n_1+i} > \delta\} \leqq 1 - \frac{p(\rho(\delta))}{2},$$

and

$$P\left\{\begin{matrix} X_{n_1+i+l} > X_{n_1+i+l-1} \mid \delta < |X_{n_1+j}| < \delta' \quad j = 1, \cdots, i \\ l = 1, \cdots, s \end{matrix}\right.$$
$$\left.\text{and } b_{n_1} = a_{t_1} \text{ and } X_{n_1+i} > \delta\right\} \leqq \left\{1 - \frac{p(\rho(\delta))}{2}\right\}^s.$$

As we pick a new a_j as soon as $(X_i - X_{i-1})(X_{i-1} - X_{i-2}) \leqq 0$, we have: Expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \quad j = 1, \cdots, i \quad \text{and} \quad b_{n_1} = a_{t_1}\}$$

under the condition that $X_{j+1} \geqq X_j > \delta$ for the first $j \geqq n_1$ with $b_j = a_{t_1+k}$, is at most

(33)
$$2 + \left(1 - \frac{p(\rho(\delta))}{2}\right) + \left(1 - \frac{p(\rho(\delta))}{2}\right)^2 + \cdots \leqq \frac{3}{p(\rho(\delta))}.$$

The case where $X_{j+1} < X_j$ at the first time that $b_j = a_{t_1+k}$ is more difficult. Let us divide the interval (δ, δ') in

$$\left[\frac{2(\delta'-\delta)}{\rho(\delta)a_{t_1+k}}\right] + 1$$

non-overlapping intervals[5] I_i with

[5] $[a]$ is the largest integer $\leqq a$.

$$\text{length } (I_t) < \frac{\rho(\delta)a_{t_1+k}}{2}\left(t = 1, 2, \cdots, \left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1\right).$$

Expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \, j = 1, \cdots, i, Z_{n_1+i} \varepsilon I_t\}$$

under the condition that $X_{j+1} < X_j$, $X_j > \delta$ for first $j \geqq n_1$, with $b_j = a_{t_1+k}$, is at most

$$1 + \left(1 - \frac{p(\rho(\delta))}{2}\right) + \left(1 - \frac{p(\rho(\delta))}{2}\right)^2 \cdots = \frac{2}{p(\rho(\delta))}.$$

This can be proved analogous to (33) using (22) and the fact that

$$\text{length } (I_t) < [\rho(\delta)a_{t_1+k}]/2.$$

As there are

$$\left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1$$

intervals I_t, expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \, j = 1, \cdots, i, \quad \text{and} \quad b_{n_1} = a_{t_1}\}$$

under the condition that $X_{j+1} < X_j$, $X_j > \delta$ for the first $j \geqq n_1$ with

$$b_j = a_{t_1+k},$$

is at most

$$\frac{2\left\{\left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1\right\}}{p(\rho(\delta))}.$$

Similar estimates are valid when $X_j < -\delta$ for the first $j \geqq n_1$ with $b_j = a_{t_1+k}$.

As $a_{t_1+k} \to 0 (k \to \infty)$, we can find a positive constant D such that

$$Er_{t_1+k} \leqq \frac{D}{a_{t_1+k}}.$$

By (31) and (32), it follows that

$$P\left\{\left|\sum_{n=n_1}^{n} (Z_{k+1} - Z_k - m_k)\right| \geqq \frac{\rho(\delta)}{4} \sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta)\right\}$$

$$\leqq \frac{CD \sum_{k=0}^{n-n_1} a_{t_1+k}}{\left\{\frac{\rho(\delta)}{4} \sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta)\right\}^2}. \tag{34}$$

As $\sum_{k=0}^{\infty} a_{t_1+k} = \infty$, the right hand side of (34) tends to zero when $n \to \infty$

and therefore (29) cannot be true and

$$P\{|X_n| \text{ converges to } X \neq 0 \quad \text{and} \quad t(n) \to \infty\} = 0.$$

Combining the remark after Lemma 6, and Lemma 7 we proved

$$P\{\liminf |X_n| = 0\} = 1. \tag{35}$$

Until now we only used that a_n tends monotonically to zero and

$$\textstyle\sum_{n=1}^{\infty} a_n = \infty,$$

but not yet $\sum_{n=1}^{\infty} a_n^2 < \infty$.

LEMMA 8. *Define*

$$s(n) = \begin{cases} 1 & \text{if} \quad T_n(X_1, \cdots, X_n)\cdot X_n > 0 \\ -1 & \text{if} \quad T_n(X_1, \cdots, X_n)\cdot X_n \leqq 0, \end{cases}$$

$$Y_n^1 = Y_n \prod_{j=1}^{n} s(j),$$

$$d(m, m-1) = 1$$

$$d(m, n) = \prod_{j=m}^{n} (1 + \beta_j) \qquad (n \geqq m),$$

$$S(m+1, n) = \sum_{j=m}^{n} d(j+1, n) b_j Y_j^1.$$

Then the conditions

$$\alpha_{m+j-1}(X_1, \cdots, X_{m+j-1}) \leqq \frac{\epsilon}{8} \qquad j = 1, \cdots, k,$$

$$d(m, \infty) \leqq \frac{3}{2},$$

$$|X_m| \leqq \frac{\epsilon}{4},$$

$$|X_{m+j}| > \frac{\epsilon}{4} \qquad j = 1, \cdots, k-1,$$

and

$$\sup_{n \geqq m} |S(m+1, n)| \leqq \frac{\epsilon}{16}$$

imply

$$|X_{m+j}| \leqq \frac{\epsilon}{2} \qquad j = 1, \cdots, k.$$

The proof follows immediately from Wolfowitz [6], p. 1154. We need the following

COROLLARY. *If $t(m)$ is so large that*

$$\alpha_{m+j-1}(X_1, \cdots, X_{m+j-1}) \leqq \frac{\epsilon}{8} \qquad j = 1, 2, \cdots$$

and if

$$d(m, \infty) \leqq \frac{3}{2},$$

then

$$P\left\{ |X_{m+j}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |X_m| \leqq \frac{\epsilon}{4}\right\}$$

$$\leqq P\left\{ \sup_{n_1, n_2 \geqq m} |S(n_1 + 1, n_2)| \geqq \frac{\epsilon}{16}\right\}$$

$$\leqq \frac{32^2}{\epsilon^2} \sum_{j=m}^{\infty} \operatorname{var} \{d(j+1, n) b_j Y_j^1\} \leqq \left(\frac{48\sigma}{\epsilon}\right)^2 \sum_{j=m}^{\infty} E b_j^2 .$$

PROOF OF THEOREM 1. In view of Lemma 3, we only have to prove

$$P\{\limsup |X_n| > 0 \quad \text{and} \quad t(n) \to \infty\} = 0.$$

By condition (13)

$$2\zeta = \min\Big(\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{0<|x_n|\leqq\tau} P\{X_{n+1} - X_n \geqq 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\},$$

$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{0<|x_n|\leqq\tau} P\{X_{n+1} - X_n < 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\}\Big) > 0$$

Take $\xi > 0$ and n_2 such that

$$\begin{aligned} &P\{X_{n+1} - X_n \geqq 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\} > \zeta > 0, \\ &P\{X_{n+1} - X_n < 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\} > \zeta > 0 \end{aligned} \tag{36}$$

for $0 < |x_n| \leqq \xi$ and $n \geqq n_2$.

Choose an $\epsilon \leqq \xi$ and t_2 such that

$$\alpha_n(X_1, \cdots, X_n) \leqq \frac{\epsilon}{8} \quad \text{when} \quad t(n) \geqq t_2,$$

and let

$$d(n_2, \infty) \leqq \frac{3}{2}.$$

Let now for some $m \geqq n_2$

$$|X_m| \leqq \frac{\epsilon}{4} \quad \text{and} \quad t(m) \geqq t_2.$$

Construct the following process

$$Z_k^{(m)} = X_k \quad \text{if} \quad k = 1, \cdots, m$$

and

$$Z_{m+i}^{(m)} = T_{m+i-1}(Z_1, \cdots, Z_{m+i-1}) + c_{m+i-1}Y_{m+i-1}(X_1, \cdots, X_{m+i-1}) (i = 1, 2, \cdots),$$

where the c's are determined in the following way:

$$c_m = b_m = a_{t(m)}$$

$$(37)\quad c_{m+i} = \begin{cases} c_{m+i-1} & \text{if } |Z_{m+j}| \leqq \frac{\epsilon}{2} \quad j = 0, 1, \cdots, i \\ & \text{and } (Z_{m+i} - Z_{m+i-1})(Z_{m+i-1} - Z_{m+i-2}) > 0 \\ a_l & \text{if } |Z_{m+j}| \leqq \frac{\epsilon}{2} \quad j = 0, 1, \cdots, i \\ & \text{and } (Z_{m+i} - Z_{m+i-1})(Z_{m+i-1} - Z_{m+i-2}) \leqq 0 \\ & \text{and } c_{m+i-1} = a_{l-1} \\ a_{t(m)+i} & \text{otherwise.} \end{cases}$$

Then Er_l = expected number of times $c_{m+j} = al$ is zero when $l < t(m)$. For $l \geqq t(m)$ it is at most

$$(38)\qquad 1 + (1 - \zeta) + (1 - \zeta)^2 \cdots = \frac{1}{\zeta}.$$

In fact from (36) and (37),

$$P\{c_{m+j} = c_{m+j-1}\} \leqq 1 - \zeta.$$

Using (38) and applying the corollary of Lemma 8 to the $Z(m)$ process, and thus replacing the b's by the c's, one finds for $m \geqq n_2$,

$$P\left\{|X_{m+j}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |X_m| \leqq \frac{\epsilon}{4}, t(m) \geqq t_2\right\}$$

$$\leqq P\left\{|Z_{m+j}^{(m)}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |Z_m^{(m)}| \leqq \frac{\epsilon}{4}, t(m) \geqq t_2\right\}$$

$$\leqq \left(\frac{48\sigma}{\epsilon}\right)^2 \sum_{n=t_2}^{\infty} \frac{a_n^2}{\zeta}.$$

Now choose $t_3 \geqq t_2$ such that

$$\left(\frac{48\sigma}{\epsilon}\right)^2 \frac{1}{\zeta} \sum_{n=t_3}^{\infty} a_n^2 \leqq \frac{\epsilon}{2},$$

and $n_3 \geqq n_2$ such that

$$P\left\{\left[\,|X_n| > \frac{\epsilon}{4} \quad \text{for all} \quad n \geqq n_3 \quad \text{or} \quad t(n_3) < t_3\right] \quad \text{and} \quad t(n) \to \infty\right\} \leqq \frac{\epsilon}{2}$$

(such an n_3 exists by (35)). Then

$$P\{\limsup |X_n| > \epsilon \quad \text{and} \quad t(n) \to \infty\}$$

$$\leqq P\left\{\left[\,|X_n| > \frac{\epsilon}{4} \quad \text{for all} \quad n \geqq n_3 \quad \text{or} \quad t(n_3) < t_3\right] \quad \text{and} \quad t(n) \to \infty\right\}$$

$$+ \sum_{m=n_3}^{\infty} P\left\{X_m \quad \text{is the first after} \quad X_{n_3-1} \quad \text{with} \quad |X_m| \leqq \frac{\epsilon}{4}\right.$$

$$\left.\text{and} \quad t(n_3) \geqq t_3 \quad \text{and} \quad \max_{k \geqq m} |Z_k(m)| > \frac{\epsilon}{2}\right\}$$

$$\leqq \frac{\epsilon}{2} + \frac{\epsilon}{2} \sum_{m=n_3}^{\infty} P\left\{X_m \quad \text{is the first after} \quad X_{n_3-1} \quad \text{with} \quad |X_m| \leqq \frac{\epsilon}{4}\right\} \leqq \epsilon.$$

As the only restriction on ϵ is $\epsilon \leqq \xi$, this proves the theorem.

3. Applications.

Accelerated Robbins-Monro procedure.

THEOREM 2. *Let X_1 and $Y(x)$ be random variables and $\{a_n\}$ a sequence of positive numbers and define*

$$X_{n+1}(\omega) = X_n(\omega) - b_n(M(X_n) - \alpha) + b_n Y(X_n).$$

The sequence $\{b_n\}$ is selected in the following way from the sequence $\{a_n\}$:

$$b_1 = a_1,$$

$$b_2 = a_2,$$

$$b_n = a_{t(n)},$$

(cf. (2) and (3)).

If $M(x)$ is a measurable function satisfying

(39) $$(x - \theta)(M(x) - \alpha) > 0 \qquad \text{for } x \neq \theta,$$

(40) $$\inf_{\delta \leqq |x-\theta| < \infty} |M(x) - \alpha| > 0 \qquad \text{for every} \quad \delta > 0,$$

(41) $|M(x) - \alpha| \leqq c + d\,|x - \theta|$ for some positive constants c and d,

and if

(42) $$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \text{and} \quad a_{n+1} \leqq a_n,$$

(43) $$E(Y(X_n) | X_1, \cdots, X_n) = 0, \qquad E(Y^2(X_n) | X_1, \cdots, X_n) \leqq \sigma^2$$

with probability 1,

$$
\begin{aligned}
&\lim_{\tau\to 0} \inf_{0<|x-\theta|\le\tau} P\{Y(x) - M(x) + \alpha \geqq 0\} > 0 \\
(44)\qquad &\lim_{\tau\to 0} \inf_{0<|x-\theta|\le\tau} P\{Y(x) - M(x) + \alpha < 0\} > 0,
\end{aligned}
$$

then

$$P\{X_n \text{ converges to } \theta\} = 1.$$

Proof. Take

$$
\alpha_n = \begin{cases} b_n(c + d\,|\,x_n - \theta\,|) & \text{for } b_n d > 1 \\ b_n c & \text{for } b_n d \leqq 1, \end{cases}
$$

$$\beta_n \equiv 0,$$

$$\gamma_n = b_n\,|\,M(x_n) - \alpha\,|$$

in Theorem 1.

The process as described in Theorem 2 gives a stochastic approximation method for the point θ which uses the number of changes in sign in

$$(X_i - X_{i-1})(X_{i-1} - X_{i-2}) \qquad i = 3, \cdots, n$$

to determine $(X_{n+1} - X_n)$. We only reduce b_n and thus the magnitude of

$$X_{n+1} - X_n$$

when the last two corrections $X_n - X_{n-1}$ and $X_{n-1} - X_{n-2}$ had different signs. As indicated in the summary this process may pull X_n to θ faster (for large $|\,X_n - \theta\,|$) than the Robbins-Monro procedure. In Theorem 2 the conditions are slightly stronger than for the Robbins-Monro process as given by Blum [1]. Blum does not need

$$a_{n+1} \leqq a_n$$

or (44) and has

$$(40a)\qquad \inf_{\delta\le|x-\theta|\le\delta'} |\,M(x) - \alpha\,| > 0 \quad \text{for every} \quad 0 < \delta \leqq \delta' < \infty$$

instead of (40).

One can easily give an example to show that we cannot replace (40) by (40a) and the following example shows that (44) cannot be dispensed with.

Example. Take

$$\theta = 0, \qquad \alpha = 0, \qquad a_n = \frac{1}{n}.$$

Let $\{x_{2n+1,0}\}(n = 0, 1, \cdots)$ be a sequence of real numbers such that

$$x_{2n+1,0} \neq x_{2m+1,0} \quad \text{for} \quad n \neq m \quad \text{and} \quad 1 \leqq x_{2n+1,0} \leqq 2.$$

Let $\{x_{2n,0}\}(n = 1, 2, \cdots)$ be a sequence of real numbers such that

$$x_{2n,0} \neq x_{2m,0} \quad \text{for} \quad n \neq m \quad \text{and} \quad -2 \leqq x_{2n,0} \leqq -1.$$

We now construct recursively sequences $\{x_{n,k}\}(k = 0, 1, \cdots)$. Put

$$Z(x) = M(x) - Y(x)$$

and

$$Z_{n,k} = Z(x_{n,k}),$$

so

$$X_{n+1} = X_n - b_n Z(X_n)$$

We start with $\{x_{1,k}\}$ by taking

$$Z_{1,0} = \begin{cases} z'_{1,0} = x_{1,0} - x_{2,0} & \text{with probability } \frac{1}{2} \\ z''_{1,0} & \text{with probability } \frac{1}{2}, \end{cases}$$

where $\frac{1}{2}x_{1,0} < z''_{1,0} < x_{1,0}$. Further take $x_{1,1} = x_{1,0} - z''_{1,0}$, and in general

$$Z_{1,k} = \begin{cases} z'_{1,k} = x_{1,k} - x_{2,0} & \text{with probability } \frac{1}{2} \\ z''_{1,k} & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$x_{1,k+1} = x_{1,k} - z''_{1,k},$$

where

$$\tfrac{1}{2}x_{1,k} < z''_{1,k} < x_{1,k}.$$

For $n > 1$ we take

$$Z_{n,0} = \begin{cases} z'_{n,0} = (n-1)(x_{n,0} - x_{n+1,0}) & \text{with probability } \dfrac{1}{2n^2} \\ z''_{n,0} & \text{with probability } 1 - \dfrac{1}{2n^2}, \end{cases}$$

$$x_{n,1} = x_{n,0} - \frac{1}{n-1} z''_{n,0},$$

where $z''_{n,0}$ is such that

$$z''_{n,0} \cdot x_{n,0} > 0 \quad \text{and} \quad \tfrac{1}{2}|x_{n,0}| < |z''_{n,0}| < |x_{n,0}|$$

and $x_{n,1}$ is not equal to any $x_{m,i}$ with $m < n$. Further for $k > 0$,

$$Z_{n,k} = \begin{cases} z'_{n,k} = n(x_{n,k} - x_{n+1,0}) & \text{with probability } \dfrac{1}{2n^2} \\ z''_{n,k} & \text{with probability } 1 - \dfrac{1}{2n^2}, \end{cases}$$

$$x_{n,k+1} = x_{n,k} - \frac{1}{n} z''_{n,k},$$

where $z''_{n,k}$ is such that

$$z''_{n,k} \cdot x_{n,k} > 0, \qquad \tfrac{1}{2} \mid x_{n,k} \mid < \mid z''_{n,k} \mid < \mid x_{n,k} \mid$$

and $x_{n,k+1}$ is not equal to any $x_{m,i}$ with $m < n$. We take $M(x_{n,k}) = EZ_{n,k}$ and

$$Y(x_{n,k}) = Z_{n,k} - M(x_{n,k}).$$

For $x \neq x_{n,k}$ for all n, k, we take $M(x)$ and $Y(x)$ in any way such that the conditions of Theorem 2, except (44), are satisfied.

Take now $X_1 = x_{1,0}$ with probability 1. By the choice of $z'_{n,k}$, we get the value $x_{n+1,0}$ as soon as $Z_{n,k}$ takes the value $z'_{n,k}$. But for every n, with probability 1, Z will take once the value $z'_{n,k}$. Therefore with probability 1, all the values $x_{n,0}$ occur in the sequence $\{X_n\}$ and thus,

$$P\{X_n \text{ converges to } 0\} = 0.$$

Accelerated Kiefer-Wolfowitz procedure.

THEOREM 3. *Let X_1 and $Y(x)$ be random variables and let $\{a_n\}$ be a sequence of positive numbers and u some positive constant and define*

$$X_{n+1}(\omega) = X_n(\omega) - b_n[M(X_n - u) - M(X_n + u)] + b_n[Y(X_n - u) - Y(X_n + u)].$$

The sequence $\{b_n\}$ is selected in the following way from the sequence $\{a_n\}$:

$$b_1 = a_1,$$
$$b_2 = a_2,$$
$$b_n = a_{s(n)},$$

(cf. (2) and (3)). *If $M(x)$ is a measurable function, satisfying*

$$(45) \qquad \begin{aligned} &\inf_{x-\theta \geq \delta} \{M(x-u) - M(x+u)\} > 0 \\ &\inf_{x-\theta \leq -\delta} \{M(x-u) - M(x+u)\} < 0 \end{aligned}$$

for every $\delta > 0$,

$$(46) \qquad | M(x-u) - M(x+u)| \leq c + d \mid x - \theta \mid$$

for some positive constants c and d, and if

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \text{and} \quad a_{n+1} \leqq a_n, \tag{47}$$

$$\begin{gathered} E(Y(X_n - u) - Y(X_n + u) \mid X_1, \cdots, X_n) = 0, \\ E((Y(X_n - u) - Y(X_n + u))^2 \mid X_1, \cdots, X_n) \leqq \sigma^2 \end{gathered} \tag{48}$$

with probability 1,

$$\begin{gathered} \lim_{\tau \to 0} \inf_{0 < |x-\theta| \leqq \tau} P\{Y(x-u) - Y(x+u) - M(x-u) + M(x+u) \\ \geqq 0\} > 0 \\ \lim_{\tau \to 0} \inf_{0 < |x-\theta| \leqq \tau} P\{Y(x-u) - Y(x+u) - M(x-u) + M(x+u) \\ < 0\} > 0, \end{gathered} \tag{49}$$

then

$$P\{X_n \text{ converges to } \theta\} = 1.$$

PROOF. Take

$$\alpha_n = \begin{cases} b_n(c + d \mid x_n - \theta \mid) & \text{for} \quad b_n d > 1 \\ b_n c & \text{for} \quad b_n d \leqq 1, \end{cases}$$

$$\beta_n \equiv 0,$$

$$\gamma_n = b_n \mid M(x_n - u) - M(x_n + u) \mid$$

in Theorem 1.

REMARK. Theorem 3 is also implied by Theorem 2. The procedure in Theorem 3 requires u to be independent of n, and therefore differs from the usual Kiefer-Wolfowitz procedure ([3]). Also condition (45) does not imply that $M(x)$ has a maximum, or if it has one, that θ is the location of the maximum. However, for every y with $|y - \theta| > u$, there exists an x with $|x - \theta| \leqq u$, such that $M(x) > M(y)$.

Acknowledgement. The author wishes to thank Professor J. Wolfowitz for suggesting the possibility of generalizing Theorem 2 and for his many other helpful and critical remarks.

REFERENCES

[1] J. R. BLUM, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386.

[2] A. Dvoretzky, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, 1956, pp. 39–55.

[3] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.

[4] M. Loève, *Probability Theory*, D. van Nostrand, 1955.

[5] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.

[6] J. Wolfowitz, "On stochastic approximation methods," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1151–1156.

TESTING THE HYPOTHESIS THAT TWO POPULATIONS DIFFER ONLY IN LOCATION[1]

By Balkrishna V. Sukhatme

Indian Council of Agricultural Research, New Delhi

0. Summary. Let $X_1, X_2, \cdots, X_n$ be n independent identically distributed random variables with cumulative distribution function $F(x - \xi)$. Let

$$\hat{\xi}(X_1, X_2, \cdots, X_n)$$

be an estimate of ξ such that $\sqrt{n}(\hat{\xi} - \xi)$ is bounded in probability. The first part of this paper (Secs. 2 through 4) is concerned with the asymptotic behavior of U-statistics modified by centering the observations at $\hat{\xi}$. A set of necessary and sufficient conditions are given under which the modified U-statistics have the same asymptotic normal distribution as the original U-statistics. These results are extended to generalized U-statistics and to functions of several generalized U-statistics. The second part gives an application of the asymptotic theory developed earlier to the problem of testing the hypothesis that two populations differ only in location.

1. Introduction. Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two independent samples of observations from populations with cumulative distribution functions $F(x - \xi)$ and $G(x - \eta) = F[(x - \eta)/\delta]$ respectively, ξ and η being the unknown location parameters and δ a scale parameter. No knowledge is assumed concerning the distribution functions F and G except that they are absolutely continuous. The problem considered in this paper is that of testing the hypothesis that the two populations differ only in location against the alternative that the Y's are more spread out than the X's and vice versa, or in symbols

$$H: \delta = 1, \quad A: \delta \neq 1. \tag{1.1}$$

From intuitive considerations and the work of Fraser [1], it seems likely that there do not exist similar tests for testing the hypothesis H, which are very satisfactory. The following simplified problem was therefore considered by the author [2]. Let the location parameters ξ and η be known, say $\xi = \eta = 0$, so that the distribution functions of X and Y differ only in the scale parameter. Then the problem considered is that of testing the hypothesis

$$H': \delta = 1, \qquad \text{i.e., } F = G,$$
$$A: \delta \neq 1, \qquad \text{i.e., } F \neq G.$$

Received May 18, 1956.

[1] This research was performed while the author was at the Statistical Laboratory, University of California, Berkeley, and was supported by the Office of Ordnance Research, U. S. Army, under contract DA-04-200-ORD-171.

Several nonparametric tests have been suggested for testing the hypothesis H', particularly by Mood [3]. The author [2] has considered some of these tests and discussed their asymptotic properties from the point of view of power considerations. These tests are based on what are known as generalised U-statistics and are reasonably efficient. But our main interest lies not in testing the hypothesis H' but H. However, once we have a class $\{W_N\}$ of tests for testing the hypothesis H', a class of tests $\{\hat{W}_N\}$ for testing the hypothesis H suggests itself. This class of tests may be obtained as follows. We obtain suitable estimates of the parameters ξ and η and then apply any of the tests of the class W_N to the deviations of the X's and the Y's from the respective estimates.

If the X's and the Y's come from normal populations, the usual test of significance for testing the hypothesis H is the variance ratio test based on the statistic

$$F = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{\sum_{i=1}^{m} (X_i - \bar{X})^2} \cdot \frac{m-1}{n-1}, \tag{1.2}$$

which is also the most commonly used statistical test for comparing sample variances. Usually, however, since little is known about the populations from which the samples are drawn, this test is used as if the assumption of normality could be ignored. It appears, however, that this is not the case.

The sensitivity to non-normality of the F-test was first pointed out by E. S. Pearson [4] whose findings were later confirmed by Geary [5] and Gayen [6]. They showed that the F-test is particularly sensitive to changes in Kurtosis from the normal theory value of zero. It is easy to see that the F-statistic when suitably normalised is asymptotically distribution free. More recently, Box and Andersen ([7] and [8]) have studied this problem in great detail and have shown on the basis of extensive sampling experiments that the F-statistic so normalized is insensitive to departures from normality.

Since the tests considered in [2] are nonparametric and reasonable for normal alternatives, it appears that they might be more efficient for non-normal alternatives and also more stable for small samples. We propose, therefore, to investigate whether such tests, after modification by the introduction of estimates of parameters are asymptotically distribution free.

This is achieved by considering the asymptotic theory of generalised U-statistics modified by the introduction of estimates of parameters, which is given in Secs. 3 and 4. In Sec. 5, it is shown that the nonparametric test proposed in [2], after modification, is asymptotically distribution free for populations with bounded and symmetric probability densities. It turns out however that even under such restrictive conditions, the nonparametric test proposed by Mood, after modification is not asymptotically distribution free. Finally, the last section considers the small sample behavior of the proposed test for some particular alternatives.

2. Some definitions and known results.

DEFINITION 2.1. Let X_{ij}, $j = 1, 2, \cdots, n_i$ for a fixed i be independent random variables identically distributed with c.d.f. $F_i(x)$ and density function $f_i(x)$. Let i run from 1 to k and $s_1 \leqq n_1$, $s_2 \leqq n_2$, $\cdots$, $s_k \leqq n_k$. Further, let

$$\varphi(u_1, \cdots, u_{s_1}; v_1, \cdots, v_{s_2}; \cdots; w_1, w_2, \cdots, w_{s_k})$$

be a function symmetric in each set of its arguments. Then the statistic

$$U_N = \binom{n_1}{s_1}^{-1}\binom{n_2}{s_2}^{-1}\cdots\binom{n_k}{s_k}^{-1}$$
$$\cdot \sum \varphi(X_{1,\alpha_1}, \cdots, X_{1,\alpha_{s_1}}; X_{2,\beta_1}, \cdots, X_{2,\beta_{s_2}}; \cdots; X_{k,\delta_1}, \cdots X_{k,\delta_{s_k}}),$$

where the summation runs over all subscripts α, β, δ such that

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_1} \leqq n_1$$
$$1 \leqq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leqq n_2$$
$$\cdots$$
$$1 \leqq \delta_1 < \delta_2 < \cdots < \delta_{s_k} \leqq n_k$$

is called a generaliesd U-statistic.

Let $\rho_1, \rho_2, \cdots, \rho_k$ be k fixed numbers such that $n_i = N\rho_i$ and $\sum_{i=1}^k \rho_i = 1$. Then Lehmann [9] has shown that $\sqrt{N}[U_N - EU_N]$ is asymptotically normally distributed with mean zero and asymptotic variance σ^2 given by

$$\sigma^2 = \frac{s_1^2}{\rho_1}\zeta_{100\cdots 0} + \frac{s_2^2}{\rho_2}\zeta_{010\cdots 0} + \cdots + \frac{s_k^2}{\rho_k}\zeta_{00\cdots 01},$$

where

$$\zeta_{00\cdots 1\cdots 0} = E\varphi_1\varphi_2 - [E\varphi_1]^2,$$

1 occurs at the ith place in $\zeta_{00\cdots 1\cdots 0}$,

$$\varphi_1 = \varphi(X_{11}, \cdots, X_{1s_1}; \cdots; X_{i1}, X_{i2}, \cdots, X_{is_i}; \cdots)$$

and φ_2 is obtained from φ_1 by replacing all the X_{jk} by X'_{jk} excepting X_{i1}, the primes denoting a new set of independent random variables. This result is a generalisation of the U-statistics considered by Hoeffding [10].

For the sake of simplicity, we shall restrict ourselves to the two sample problem only. The extension to k samples is straight forward.

DEFINITION 2.2. As before, let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be two independent samples drawn from populations with c.d.f.'s $F(x - \xi)$ and $G(x - \eta)$ respectively. Further, let

$\hat{\xi}(X_1, \cdots, X_m)$ and $\hat{\eta}(Y_1, \cdots, Y_n)$ be estimates of ξ

and η, the two location parameters. Then the generalised U-statistic with the observations centered at the respective location parameters and the modified generalised U-statistic for the two sample problem are respectively,

$$U_N = \binom{m}{s_1}^{-1} \binom{n}{s_2}^{-1} \sum_{\alpha,\beta} \varphi(X_{\alpha_1} - \xi, \cdots, X_{\alpha_{s_1}} - \xi; Y_{\beta_1} - \eta, \cdots, Y_{\beta_{s_2}} - \eta),$$

$$\hat{U}_N = \binom{m}{s_1}^{-1} \binom{n}{s_2}^{-1} \sum_{\alpha,\beta} \varphi(X_{\alpha_1} - \hat{\xi}, \cdots, X_{\alpha_{s_1}} - \hat{\xi}; Y_{\beta_1} - \hat{\eta}, \cdots, Y_{\beta_{s_2}} - \hat{\eta}),$$

where $\varphi(u_1, \cdots, u_{s_1}; v_1, \cdots, v_{s_2})$ is a function symmetric in u and in v and the summation runs over all subscripts α, β such that

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_1} \leqq m,$$

$$1 \leqq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leqq n.$$

DEFINITION 2.3. Let $\hat{W}_N$ be a test based on the statistic $\hat{U}_N$. If the asymptotic distribution of $\hat{U}_N$ is independent of the original populations from which the X's and the Y's are drawn under the null hypothesis, the test $\hat{W}_N$ will be said to be asymptotically distribution free.

Finally we define a quantity L_N required in the study of the asymptotic behavior of modified generalised U statistics.

DEFINITION 2.4.

$$L_N = \binom{m}{s_1}^{-1} \binom{n}{s_2}^{-1} \cdot \sum_{\alpha,\beta} [\varphi(X_{\alpha_1} - \hat{\xi}, \cdots, X_{\alpha_{s_1}} - \hat{\xi}; Y_{\beta_1} - \hat{\eta}, \cdots, Y_{\beta_{s_2}} - \hat{\eta}) - A(\hat{\xi} - \xi, \hat{\eta} - \eta)],$$

where

$$A(t_1 - \xi, t_2 - \eta) = E\varphi(X_1 - t_1, \cdots, X_{s_1} - t_1; Y_1 - t_2, \cdots, Y_{s_2} - t_2),$$

expectation being taken with respect to all the X's and the Y's.

3. The limiting distribution of L_N. In this section, we will prove theorems, giving the conditions under which L_N and U_N have the same asymptotic normal distribution. We will start with one sample problem and then extend the result to two samples. In what follows, we write $\mathfrak{L}(X_n) \to \mathfrak{L}(X)$ (read: the distribution law of X_n converges to the distribution law of X), or $\lim_{n\to\infty} \mathfrak{L}(X_n) = \mathfrak{L}(X)$ if $F_n(a) \to F(a)$ at every point a of continuity of F where F_n and F are the c.d.f.'s of X_n and X, respectively.

THEOREM 3.1. *Let $X_1, X_2, \cdots, X_n$ be n independent identically distributed* *random variables with c.d.f.* $F(x - \xi)$. *Let* $\varphi(u_1, u_2, \cdots, u_s)$ *with* $s \leqq n$ *be a real valued symmetric function of its arguments such that if*

$$W(x_1, x_2, \cdots, x_s, t) = \varphi(x_1 - t, \cdots, x_s - t) - A(t - \xi), \tag{3.0}$$

where $A(t - \xi) = E\varphi(X_1 - t, \cdots, X_s - t)$, *the following conditions are satisfied.*

$$\text{(B}_1\text{)} \quad \begin{aligned} &|W(x_1, x_2, \cdots, x_s, t)| \leqq M_1, \text{ and } E|W(X_1, \cdots, X_s; t + h) \\ &\qquad - W(X_1, \cdots, X_s; t)| \leqq M_2 h, \; M_1 \text{ and } M_2 \text{ being fixed constants.} \end{aligned}$$

There exists a sequence $\{t_j\}$ *such that for each set of* x's

$$\text{(B}_2\text{)} \quad \begin{aligned} &\sup_{0 \leqq t_j \leqq k} | W(x_1, \cdots, x_s, t_j) - W(x_1, \cdots, x_s, 0) | \\ &\qquad = \sup_{0 \leqq t \leqq k} | W(x_1, \cdots, x_s, t) - W(x_1, \cdots, x_s, 0) |. \end{aligned}$$

Further, let $\hat{\xi}(X_1, X_2, \cdots, X_n)$ *be an estimate of* ξ *such that given* $\Sigma_1 > 0$, *there exists a number* b *such that for* n *sufficiently large*

$$P\left\{| \hat{\xi} - \xi | \geqq \frac{b}{\sqrt{n}}\right\} \leqq \Sigma_1. \tag{3.1}$$

Define

$$U_n = \binom{n}{s}^{-1} \sum \varphi(X_{\alpha_1} - \xi, \cdots, X_{\alpha_s} - \xi), \tag{3.2}$$

the summation being taken over all subscripts α *such that*

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_s \leqq n$$

and

$$L_n = \binom{n}{s}^{-1} \sum [\varphi(X_{\alpha_1} - \hat{\xi}, \cdots, X_{\alpha_s} - \hat{\xi}) - A(\hat{\xi} - \xi)].$$

Then

$$\begin{aligned} \lim_{n \to \infty} \mathcal{L}(\sqrt{n} L_n) &= \lim_{n \to \infty} \mathcal{L}(\sqrt{n}[U_n - EU_n]) \\ &= N(0, s^2\zeta_1), \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} \zeta_1 &= E\varphi_1^2(X_1 - \xi) - E^2\varphi(X_1 - \xi, \cdots, X_s - \xi), \\ \varphi_1(x_1 - \xi) &= E\varphi(x_1 - \xi, X_2 - \xi, \cdots, X_s - \xi). \end{aligned} \tag{3.4}$$

PROOF. For the sake of simplicity we may, without loss of generality, assume $\xi = 0$. However, before we proceed further, we shall first prove the following lemma, which we shall use in the proof of the theorem.

LEMMA 3.2. *Let*

$$H_{r,n}(x_1, x_2, \cdots, x_s, t) = \sup_{\frac{r\delta}{\sqrt{n}} \leq z \leq t} \left| W(x_1, \cdots, x_s, z) - W\left(x_1, \cdots, x_s, \frac{r\delta}{\sqrt{n}}\right) \right| \tag{3.5}$$

and

$$S_{r,n}(t) = \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left[W\left(x_{\alpha_1}, \cdots, x_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(x_{\alpha_1}, \cdots, x_{\alpha_s}, \frac{\delta r}{\sqrt{n}}\right)\right]. \tag{3.6}$$

Then, if $r\delta \leq t \leq (r+1)\delta$ and n is sufficiently large,

$$\text{(i)} \quad EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \leq \frac{M_2\delta}{\sqrt{n}}; \tag{3.7}$$

$$\begin{aligned} \text{(ii)} \quad E\Bigg\{\Bigg[&H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \\ &- EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\Bigg] \\ &\cdot \Bigg[H_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \\ &- EH_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\Bigg]\Bigg\} \to 0 \text{ as } n \to \infty, \end{aligned} \tag{3.8}$$

where

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_s \leq n,$$

$$1 \leq \beta_1 < \beta_2 < \cdots < \beta_s \leq n;$$

$$\text{(iii)} \quad E \mid S_{r,n}(t) \mid^2 \leq \frac{d(t - r\delta)}{\sqrt{n}}, \tag{3.9}$$

where d is a fixed constant and higher powers of $1/\sqrt{n}$ are neglected.

PROOF. (i) and (ii) are easily obtained as consequences of conditions (B_1)

and (B_2) of Theorem 3.1. To prove (iii) we have

$$E \mid S_{r,n}(t) \mid^2$$

$$= n \binom{n}{s}^{-2} \sum_{\alpha,\beta} E\left\{\left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right.$$

$$\left.\cdot\left[W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right\}.$$

Consider a typical term; with c integers common to the two terms. We then have

$$E\left\{\left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right.$$

$$\left.\cdot\left[W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right\}$$

$$\leqq E\left|\left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right.$$

$$\left.\cdot\left[W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right|$$

$$\leqq 2M_1 E\left| W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right|$$

$$\leqq 2M_1 M_2 \frac{(t - r\delta)}{\sqrt{n}}.$$

The total contribution of such terms to

$$E|S_{r,n}(t)|^2 \leqq n \binom{n}{s}^{-2} \binom{n}{2s - c} \cdot A(t - r\delta) / \sqrt{n},$$

A being some fixed constant. It follows that

$$E \mid S_{r,n}(t) \mid^2 \sim \frac{1}{n^{c-\frac{1}{2}}} (t - r\delta).$$

When $c = 0$, the expectation of the product is zero. Retaining only powers of $1/\sqrt{n}$, the result now follows. Q.E.D.

Proof of Theorem 3.1. Let

$$S_n^{(t)} = \sqrt{n} \binom{n}{s}^{-1} \sum_{\alpha} \left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, 0\right)\right].$$

Then it is easily seen that

$$S_n(t) = S_{r,n}(t) + S_{0,n}(r\delta).$$

Let $\epsilon > 0$, $\delta = \epsilon/2M_2$, and t be such that $r\delta \leqq t \leqq (r+1)\delta$. Then it is seen that

$$
\begin{aligned}
| S_{r,n}(t) | &\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) \\
&\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \\
&\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right. \\
&\qquad \left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right] \\
&\qquad + \sqrt{n}EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \\
&\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left| H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right. \\
&\qquad \left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right| + M_2\delta \\
&= D_1 + M_2\delta,
\end{aligned}
$$

where

$$
\begin{aligned}
D_1 = \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} &\left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right. \\
&\qquad \left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right].
\end{aligned}
$$

Now

$$
\begin{aligned}
ED_1^2 = n\binom{n}{s}^{-2} \sum_{\alpha,\beta} E&\left\{\left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right.\right. \\
&\qquad \left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right] \\
&\cdot\left.\left[H_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{\delta(r+1)}{\sqrt{n}}\right)\right]\right\},
\end{aligned}
$$

the summation having the same meaning as before. Considering again a typical term with c integers common to the two terms, we find that the total contribution of such terms to ED_1^2 is

$$\leq n\binom{n}{s}^{-2}\binom{n}{2s-c}E\left|\left[H_{r,n}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right.\right.$$
$$\left.-EH_{r,n}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]$$
$$\left.\cdot\left[H_{r,n}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right)-EH_{r,n}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]\right|$$
$$\sim\frac{A}{n^{c-1}}E\left|\left[H_{r,n}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)-EH_{r,n}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]\right.$$
$$\left.\cdot\left[H_{r,n}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right)-EH_{r,n}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]\right|,$$

which tends to zero by Lemma 3.2 for $c \geq 1$. When $c = 0$, the expectation of the product is identically equal to zero.

Hence $ED_1^2 \to 0$ as $n \to \infty$. It follows that

$$P\left\{\left|\sup_{r\delta\leq t\leq(r+1)\delta} S_{r,n}(t)\right| > \epsilon\right\} \to 0.$$

for every r

Also $E|S_{0,n}(r\delta)|^2 \leq 2M_1M_2r\delta / \sqrt{n} \to 0$ as $n \to \infty$, therefore,

$$S_{0,n}(r\delta) \xrightarrow{P} 0.$$

It follows that $\sup_{t\epsilon C} |S_n(t)| \to 0$, C being some bounded set. Hence,

$$S_n(\sqrt{n}\,\xi) \xrightarrow{P} 0,$$

that is,

$$\sqrt{n}\,L_n - \sqrt{n}\,[U_n - EU_n] \xrightarrow{P} 0,$$

therefore,

$$\lim_{n\to\infty}\mathfrak{L}(\sqrt{n}\,L_n) = \lim_{n\to\infty}\mathfrak{L}(\sqrt{n}\,[U_n - EU_n]).$$

But by Hoeffding's Theorem 7.1, page 305 of [10], U_n is asymptotically normally distributed, whence the required result follows. Q.E.D.

We complete this section by stating without proof the generalization of the above result to the two sample problem. The proof goes more or less along the same lines as that of Theorem 3.1.

THEOREM 3.3. *Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two independent samples drawn from populations with c.d.f.'s $F(x - \xi)$ and $G(x - \eta)$ respectively. Further, let $\varphi(u_1, \cdots, u_{s_1}; v_1, \cdots, v_{s_2})$ with $s_1 \leq m$ and $s_2 \leq n$ be a real-valued* *function symmetric in u and in v separately such that if*

$$W(x_1, x_2, \cdots, x_{s_1}, y_1, \cdots, y_{s_2}, t_1, t_2) = \varphi(x_1 - t_1, \cdots, x_{s_1} - t_1; y_1 - t_2, \cdots, y_{s_2} - t_2) - A(t_1 - \xi, t_2 - \eta), \tag{3.10}$$

the following conditions are satisfied:

$$|W(x_1, x_2, \cdots, x_{s_1}, y_1, y_2, \cdots, y_{s_2}, t_1, t_2| \leqq M_{11}$$

$$\text{(B}_3\text{)}\quad \begin{aligned} &E|W(X_1, \cdots, X_{s_1}, Y_1, \cdots, Y_{s_2}, t_1 + h, t_2) \\ &\qquad - W(X_1, \cdots, X_{s_1}, Y_1, \cdots Y_{s_2}, t_1, t_2)| \leqq M_{21}h \\ &E|W(X_1, \cdots X_{s_1}, Y_1, \cdots, Y_{s_2}, t_1, t_2 + k) \\ &\qquad - W(X_1, \cdots, X_{s_1}, Y_1, \cdots Y_{s_2}, t_1, t_2)| \leqq M_{22}k, \end{aligned}$$

where M_{11}, M_{21} and M_{22} are certain fixed constants.

There exist sequences $\{t_j\}$ and $\{l_j\}$ such that for every set of x's and y's,

$$\text{(B}_4\text{)}\quad \begin{aligned} &\sup_{\substack{0 \leqq t_j \leqq k_1 \\ 0 \leqq l_j \leqq k_2}} |W(x_1, \cdots, x_{s_1}; y_1, \cdots, y_{s_2}, t_j, l_j) \\ &\qquad - W(x_1, x_{s_1}; y_1, \cdots, y_{s_2}, 0, 0)| \\ &= \sup_{\substack{0 \leqq t \leqq k_1 \\ 0 \leqq l \leqq k_2}} |W(x_1, \cdots, x_{s_1}; y_1, \cdots, y_{s_2}, t, l) \\ &\qquad - W(x_1, \cdots, x_{s_1}; y_1, \cdots, y_{s_2}, 0, 0)|. \end{aligned}$$

Further, let $\hat{\xi}(X_1, \cdots X_m)$ and $\hat{\eta}(Y_1, \cdots, Y_n)$ be estimates of ξ and η respectively such that given $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there exist numbers b_1 and b_2 such that for m and n sufficiently large

$$P\left\{|\hat{\xi} - \xi| \geqq \frac{b_1}{\sqrt{n}}\right\} \leqq \epsilon_1, \tag{3.11}$$

$$P\left\{|\hat{\eta} - \eta| \geqq \frac{b_2}{\sqrt{n}}\right\} \leqq \epsilon_2. \tag{3.12}$$

Define

$$U_N = \binom{m}{s_1}^{-1}\binom{n}{s_2}^{-1}\sum_{\alpha,\beta}\varphi(X_{\alpha_1} - \xi, \cdots, X_{\alpha_{s_1}} - \xi; Y_{\beta_1} - \eta, \cdots, Y_{\beta_{s_2}} - \eta), \tag{3.13}$$

the summation being taken over all subscripts α, β such that

$$\begin{aligned} 1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_1} \leqq m, \\ 1 \leqq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leqq n. \end{aligned} \tag{3.14}$$

Then if $m = N\rho$ *and* $n = N(1 - \rho)$,

$$\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}\, L_N) = \lim_{N\to\infty} \mathfrak{L}(\sqrt{N}\,(U_N - EU_N)) = N(0, \sigma^2), \tag{3.15}$$

where σ^2 *is the asymptotic variance of* U_N *and is given by*

$$\sigma^2 = \frac{s_1^2}{\rho}\zeta_{10} + \frac{s_2^2}{1-\rho}\zeta_{01}, \tag{3.16}$$

where ζ_{10} *and* ζ_{01} *have the same meaning as in* (2.1).

4. The asymptotic distribution of modified generalised U-statistics. We are now in a position to consider the statistic $\hat{U}_N$ and obtain conditions under which it has the same asymptotic normal distribution as the statistic U_N. This result is contained in Theorem 4.1.

THEOREM 4.1. *If in addition to the conditions of Theorem* 3.1,

(i) $\sqrt{n}(\hat{\xi} - \xi)$ *has a limiting distribution*

and

(ii) $A(t) = E[\varphi(X_1 - t, \cdots, X_s - t) \mid \xi = 0]$ *has a derivative continuous in the neighbourhood of the origin, then*

(a) *If* $A'(0) = 0$, *where* $A'(t) = \dfrac{d}{dt}A(t)$,

$$\lim_{n\to\infty} \mathfrak{L}(\sqrt{n}\,[\hat{U}_n - EU_n]) = \lim_{n\to\infty} \mathfrak{L}(\sqrt{n}\,[U_n - EU_n]) = N(0, s^2\zeta_1).$$

(b) *If* $A'(0) \neq 0$, $\hat{\xi}$ *is asymptotically normally distributed and the joint distribution of* $\hat{\xi}$ *and* U_n *is asymptotically normal, then* $\sqrt{n}(\hat{U}_n - EU_n)$ *is asymptotically normally distributed.*

PROOF. We have

$$\sqrt{n}[\hat{U}_n - EU_n] = \sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)] + \sqrt{n}[A(\hat{\xi} - \xi) - EU_n].$$

But $A(\hat{\xi} - \xi) = A(0) + (\hat{\xi} - \xi)A'(h)$ where $h = \Delta(\hat{\xi} - \xi)$, $|\Delta| < 1$. Therefore

$$\sqrt{n}[\hat{U}_n - EU_n] = \sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)] + \sqrt{n}(\hat{\xi} - \xi)\cdot A'(h).$$

Since $\sqrt{n}(\hat{\xi} - \xi)$ has a limiting distribution and $A'(0) = 0$, it follows from the continuity considerations and Slutsky's theorem that $\sqrt{n}[\hat{U}_n - EU_n]$ and $\sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)]$ have the same asymptotic distribution. But by Theorem 3.1, $\sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)]$ and $\sqrt{n}[U_n - EU_n]$ have the same asymptotic normal distribution. It follows that $\sqrt{n}[\hat{U}_n - EU_n]$ and $\sqrt{n}[U_n - EU_n]$ have the same asymptotic normal distribution. This proves (a).

To prove (b), it is sufficient to remark that because of Theorem 3.1 and Slutsky's Theorem, the joint distribution of $\sqrt{n}(\hat{\xi} - \xi)$ and $\sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)]$ is asymptotically normal. Q.E.D.

In the preceding theorem we make the following observations.

(1) If $A'(0) = 0$, then $\sigma^2(\hat{U}_n) = \sigma^2(U_n)$.

(2) If $A'(0) \neq 0$, then $\sigma^2(\hat{U}_n) = \sigma^2(U_n)$, if and only if

$$A'(0) = \frac{-2\sigma(U_n, \hat{\xi})}{\sigma^2(\hat{\xi})},$$

where $\sigma(U_n, \hat{\xi})$ is the asymptotic covariance between U_n and $\hat{\xi}$ and $\sigma^2(\hat{\xi})$ is the asymptotic variance of $\hat{\xi}$.

For the sake of simplicity we will now consider the special case when $s = 1$, $\hat{\xi}$ is the sample median and $f(x)$ is symmetric about the median which may be taken to be the origin.

Now

$$\begin{aligned} A(t) &= E\varphi(X - t) \\ &= \int \varphi(x - t)f(x)\,dx \\ &= \int \varphi(y)f(y + t)\,dy. \end{aligned}$$

If there exists an integrable function $g(y)$ such that

$$\left|\frac{f(y + t) - f(y + t_0)}{t - t_0}\right| \leqq g(y) \tag{4.1}$$

and the derivative of f exists almost everywhere except for a set of measure zero, then

$$A'(0) = \int \varphi(y)f'(y)\,dy. \tag{4.2}$$

Also it has been shown in [11] that the joint distribution of U_n and ξ is asymptotically normal and that

$$\sigma^2(\hat{\xi}) = \frac{1}{4nf^2(0)} \tag{4.3}$$

and

$$\sigma(U_n, \hat{\xi}) = \frac{1}{2nf(0)} \int_0^\infty [\varphi(x) - \varphi(-x)]f(x)\,dx. \tag{4.4}$$

Hence $\sigma^2(\hat{U}_n) = \sigma^2(U_n)$ if and only if

$$\int_0^\infty \left[4f(0) + \frac{f'(x)}{f(x)}\right][\varphi(x) - \varphi(-x)]f(x)\,dx \underset{f}{\equiv} 0. \tag{4.5}$$

We will now show that the condition (4.5) implies that $\varphi(x) - \varphi(-x) = 0$ almost everywhere. To show this, it is enough to consider the subfamily of probability densities given by

$$f(x, \theta) = \frac{1}{2\theta} e^{-|x|/\theta}. \tag{4.6}$$

We observe that the derivative of f exists everywhere except at the origin. Also, we have

$$\left| \frac{e^{-|x+h|/\theta} - e^{-|x|/\theta}}{h/\theta} \right| \leqq ce^{-|x|/\theta}, \tag{4.7}$$

for h sufficiently small, c being a fixed constant. Condition (4.1) is thus satisfied for the family of distributions (4.6). On substitution, condition (4.5) becomes

$$\int_0^\infty e^{-x/\theta}[\varphi(x) - \varphi(-x)]\, dx \underset{\theta}{\equiv} 0, \tag{4.8}$$

whence it follows from the unicity of the unilateral Laplace transform that $\varphi(x) - \varphi(-x) = 0$ almost everywhere, in which case $A'(0) = 0$ and condition (2) reduces to condition (1).

It is now clear that $A'(0) = 0$ is a necessary and sufficient condition that $\hat{U}_n$ and U_n have the same asymptotic normal distribution.

We will now extend the results of Theorem 4.1 to the two-sample problem

THEOREM 4.2. *If in addition to the conditions of Theorem 3.3,*

(i) $\sqrt{N}(\hat{\xi} - \xi)$ *and* $\sqrt{N}(\hat{\eta} - \eta)$ *have limiting distributions*

and

(ii) $A(t_1, t_2) = E[\varphi(X_1 - t_1, \cdots, X_{s_1} - t_1, Y_1 - t_2, \cdots, Y_{s_2} - t_2) \mid \xi = \eta = 0]$

possesses first order partial derivatives continuous in the neighborhood of the origin, then

(a) *If*

$$\left. \frac{\partial A(t_1, t_2)}{\partial t_1} \right|_{t_1 = t_2 = 0} = \left. \frac{\partial A(t_1, t_2)}{\partial t_2} \right|_{t_1 = t_2 = 0} = 0,$$

$$\begin{aligned} \lim_{N \to \infty} \mathcal{L}(\sqrt{N}(\hat{U}_N - EU_N)) &= \lim_{N \to \infty} \mathcal{L}(\sqrt{N}[U_N - EU_N]) \\ &= N(0, \sigma^2), \end{aligned}$$

where σ^2 *is the asymptotic variance of* U_N.

(b) If the above condition is not satisfied, $\hat{\xi}$ and $\hat{\eta}$ are asymptotically normally distributed and the joint distribution of $\hat{\xi}$, $\hat{\eta}$ and the U statistic is asymptotically normal, then $\sqrt{N}[\hat{U}_N - EU_N]$ is asymptotically normally distributed.

PROOF. The proof of this theorem goes in exactly the same lines as that of Theorem 4.1 and is fairly obvious. Q.E.D.

It may be remarked here that the results of Secs. 3 and 4 can be extended to random vectors as also to functions of several U-statistics. The proof follows in exactly the same way as the theorem on the asymptotic distribution of a function of moments follows from the fact of their asymptotic normality [12]. We shall content ourselves by stating an analogue of Theorem 4.2 as applied to several U-statistics.

THEOREM 4.3. *With reference to the two sample problem, let*

$$\varphi(u_1, u_2, \cdots, u_{s_1(\gamma)}; \text{for}, v_1, v_2, \cdots, v_{s_2(\gamma)}), \qquad \gamma = 1, \cdots, g,$$

with $s_1(\gamma) \leq m$ and $s_2(\gamma) \leq n$ be g real valued functions symmetric in u and in v. Further, let

$$W^{(\gamma)}(x_{\alpha_1}, \cdots, x_{\alpha_{s_1(\gamma)}}; y_{\beta_1}, \cdots, y_{\beta_{s_2(\gamma)}}, t_1, t_2)$$
$$= \varphi^{(\gamma)}(x_{\alpha_1}, \cdots, x_{\alpha_{s_1(\gamma)}}; y_{\beta_1}, \cdots, y_{\beta_{s_2(\gamma)}}) - A^{(\gamma)}(t_1, t_2),$$

where

$$A^{(\gamma)}(t_1, t_2) = E[\varphi^{(\gamma)}(X_{\alpha_1} - t_1, \cdots, X_{\alpha_{s_1(\gamma)}} - t_1;$$
$$Y_{\beta_1} - t_2, \cdots, Y_{\beta_{s_2(\gamma)}} - t_2) \mid \xi = \eta = 0]$$

possess partial derivatives continuous in the neighborhood of the origin and $W^{(\gamma)}$ satisfy the conditions (B_3) *and* (B_4) *of Theorem* 3.3 *for $\gamma = 1, \cdots, g$. Also let $\sqrt{N}(\hat{\xi} - \xi)$ and $\sqrt{N}(\hat{\eta} - \eta)$ have limiting distributions where the estimates $\hat{\xi}$ and $\hat{\eta}$ satisfy the conditions* (3.11) *and* (3.12) *of Theorem* 3.3. *Define*

$$U_N^{(\gamma)} = \binom{m}{s_1(\gamma)}^{-1} \binom{n}{s_2(\gamma)}^{-1}$$
$$\cdot \sum_{\alpha,\beta} \varphi^{(\gamma)}(X_{\alpha_1} - \xi, \cdots, X_{\alpha_{s_1(\gamma)}} - \xi; Y_{\beta_1} - \eta, \cdots, Y_{\beta_{s_2(\gamma)}} - \eta),$$

the summation having the same meaning as before. Then

(i) *a necessary and sufficient condition that the joint asymptotic distribution of*

$$\sqrt{N}(\hat{U}_N^{(1)} - EU_N^{(1)}), \quad \cdots, \quad \sqrt{N}(\hat{U}_N^{(g)} - EU_N^{(g)})$$

be the same as the joint asymptotic distribution of

$$\sqrt{N}(U_N^{(1)} - EU_N^{(1)}), \quad \cdots, \quad \sqrt{N}(U_N^{(g)} - EU_N^{(g)})$$

is that

$$\left.\frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \left.\frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_2}\right|_{t_1 = t_2 = 0} = 0$$

for $\gamma = 1, 2, \cdots, g$.

(ii) *A necessary and sufficient condition that the asymptotic distribution of $\sqrt{N}\sum_{\gamma=1}^{g} C_\gamma[\hat{U}_N^{(\gamma)} - EU_N^{(\gamma)}]$ be the same as the asymptotic distribution of*

$$\sqrt{N}\sum_{\gamma=1}^{g} C_\gamma(U_N^{(\gamma)} - EU_N^{(\gamma)})$$

is that

$$\sum_{\gamma=1}^{g} C_\gamma \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=0} = 0$$

and

$$\sum_{\gamma=1}^{g} C_\gamma \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_2}\bigg|_{t_1=t_2=0} = 0.$$

5. Consequences of Theorem 4.2. In this section, we will consider some of the tests of a class $\{\hat{W}_N\}$ based on a class of statistics $\{\hat{U}_N\}$ for testing the hypothesis that two populations differ only in location and investigate whether they are asymptotically distribution free.

Consider first the test statistic T proposed in [2] based on a sample of m X's and n Y's. The test statistics may be defined as

$$T = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} K(x_i, Y_j), \tag{5.1}$$

where

$$K(X, Y) = 1 \quad \text{if} \begin{cases} \text{either } 0 < X < Y, \\ \text{or} \quad Y < X < 0, \end{cases} \tag{5.2}$$
$$= 0 \quad \text{otherwise.}$$

A corresponding modified test is then based on the statistic

$$\hat{T} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} K(X_i - \tilde{X}, Y_j - \tilde{Y}), \tag{5.3}$$

$\tilde{X}$ and $\tilde{Y}$ being the sample medians. Let $\xi = \eta = 0$. We then have

$$\begin{aligned} A(t_1, t_2) &= EK(X - t_1, Y - t_2) \\ &= \int_0^\infty [1 - G(x + t_2)]\, dF(x + t_1) + \int_{-\infty}^0 G(x + t_2)\, dF(x + t_1). \end{aligned} \tag{5.4}$$

Also $W(x, y, t_1, t_2) = K(x - t_1, y - t_2) - A(t_1, t_2)$. It can then be shown that

$$\begin{aligned} E\,|\,W(X, Y, t_1, 0) &- W(X, Y, 0, 0)| \\ &\leqq 3 \int_{-\infty}^0 |\,F(x + t_1) - F(x)\,|\, dG(x) + 2\,|\,F(t_1) - F(0)\,| \\ &\leqq 5at_1 \end{aligned}$$

if the distribution function F has a derivative F' bounded in absolute value by a. Similarly, it can be shown that

$$E|W(X, Y, 0, t_2) - W(X, Y, 0, 0)| \leqq 5bt_2,$$

provided the distribution function G has a derivative G' bounded in absolute value by b.

Hence the condition (B_3) of Theorem 3.3 is satisfied. Observing that K can be expressed as a difference of two monotone functions, it is easy to see that condition (B_4) is also satisfied. Again, we have

$$\frac{\partial A(t_1, t_2)}{\partial t_1} = -f(t_1)[2G(t_2) - 1] + \int_0^{\infty} f(x + t_1)\, dG(x + t_2) - \int_{-\infty}^{0} f(x + t_1)\, dG(x + t_2),$$

$$\frac{\partial A(t_1, t_2)}{\partial t_2} = -\int_0^{\infty} g(x + t_2)\, dF(x + t_1) + \int_{-\infty}^{0} g(x + t_2)\, dF(x + t_1).$$

Clearly,

$$\left.\frac{\partial A(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \left.\frac{\partial A(t_1, t_2)}{\partial t_2}\right|_{t_1 = t_2 = 0} = 0$$

if $f(x)$ and $g(x)$ are symmetric about the origin. Conditions (a) of Theorem 4.2 are satisfied. Hence $\hat{T}$ has the same asymptotic normal distribution as the statistic T. This consequence is stated in Theorem 5.1.

THEOREM 5.1. *If the X's and the Y's are distributed symmetrically about the respective medians and have bounded density functions, the test of the hypothesis H based on the statistic $\hat{T}$ is asymptotically distribution free.*

Consider now the test statistic suggested by Mood [3]. The test statistic may be defined as

$$M = \sum_{i=1}^{n} \left(r_i - \frac{m + n + 1}{2}\right)^2, \tag{5.5}$$

where r_i is the rank of Y_i in the combined sample of $(m + n)$ observations. Noting that

$$r_i = 1 + \sum_{j=1}^{m} \varphi(X_j, Y_i) + \sum_{k=1}^{n} \varphi(Y_k, Y_i), \tag{5.6}$$

where

$$\begin{aligned} \varphi(u, v) &= 1 \quad \text{if} \quad u < v, \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

it is easy to see that if $m + n = N$, and

$$\begin{aligned} \psi(u, v, w) &= 1 \quad \text{if} \quad u < w \quad \text{and} \quad v < w, \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

$$\frac{M}{N^3} = C_1 U_N^{(1)} + C_2 U_N^{(2)} + C_3 U_N^{(3)} + P\left(\frac{1}{N}\right), \tag{5.7}$$

where, C_1, C_2, C_3 are certain known fixed constants, $P(1/N)$ is a third-degree polynomial in $1/N$ and

$$U_N^{(1)} = \binom{m}{2}^{-1}\binom{n}{1}^{-1}\sum_i^n\sum_{j\neq k}^m \psi(X_j, X_k, Y_i),$$

(5.8)
$$U_N^{(2)} = \binom{m}{1}^{-1}\binom{n}{2}^{-1}\sum_j^m\sum_{k\neq i}^n \psi(X_j, Y_k, Y_i),$$

$$U_N^{(3)} = \binom{m}{1}^{-1}\binom{n}{1}^{-1}\sum_i\sum_j \varphi(X_j, Y_i)$$

are three generalised U-statistics so that

(5.9)
$$\frac{\hat{M}}{N^3} = C_1\hat{U}_N^{(1)} + C_2\hat{U}_N^{(2)} + C_3\hat{U}_N^{(3)} + P\left(\frac{1}{N}\right),$$

where $\hat{U}_N^{(i)}$ is obtained from $U_N^{(i)}$ by centering the observations at the respective sample medians. Consider the statistic $\hat{U}_N^{(3)}$. We have,

$$A^{(3)}(t_1, t_2) = E\varphi(X - t_1, Y - t_2)$$
$$= \int F(x + t_1)\, dG(x + t_2),$$
$$W^{(3)}(x, y, t_1, t_2) = \varphi(x - t_1, y - t_2) - A^{(3)}(t_1, t_2).$$

It can then be shown that

$$E|W^{(3)}(X, Y, t_1, 0) - W^{(3)}(X, Y, 0, 0)| \leqq 2at_1$$

and

$$E|W^{(3)}(X, Y, 0, t_2) - W^{(3)}(X, Y, 0, 0)| \leqq 2bt_2 .$$

Condition (B_3) of Theorem 3.3 is thus satisfied. Exactly in the same manner, it can be shown that the condition (B_3) is also satisfied by the statistics $\hat{U}_N^{(1)}$ and $\hat{U}_N^{(2)}$. Condition (B_4) is also easily seen to be satisfied. Also we have

$$A^{(2)}(t_1, t_2) = E\psi(X_i - t_1, Y_j - t_2, Y_k - t_2)$$
$$= \int F(x + t_1)G(x + t_2)\, dG(x + t_2),$$
$$A^{(1)}(t_1, t_2) = E\psi(X_i - t_1, X_j - t_1, Y_k - t_2)$$
$$= \int F^2(x + t_1)\, dG(x + t_2),$$

$$\left.\frac{\partial A^{(1)}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2=0} = 2\int F(x)f(x)g(x)\, dx,$$
$$\left.\frac{\partial A^{(2)}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2=0} = \int G(x)f(x)g(x)\, dx,$$
$$\left.\frac{\partial A^{(3)}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2=0} = \int f(x)g(x)\, dx.$$

Clearly,

$$\sum_{\gamma=1}^{3} C_\gamma \left. \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_1} \right|_{t_1=t_2=0} \neq 0.$$

Similarly, it is easy to see that

$$\sum_{\gamma=1}^{3} C_\gamma \left. \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_2} \right|_{t_1=t_2=0} \neq 0.$$

Hence, it follows as a consequence of Theorem 4.3 that the statistics $\hat{M}$ and M do not have the same asymptotic normal distribution. It follows that the test based on the statistic $\hat{M}$ is not asymptotically distribution free.

6. Small sample behavior of the proposed test. It was shown in the previous section that the test statistic $\hat{T}$ is asymptotically distribution free. We will now give some idea regarding the small sample behavior of this test by considering the simplest possible case, namely $m = n = 3$. The computations involved even in this relatively simple case are very extensive. We will consider the one sided test of the hypothesis

$$H:\delta = 1,$$

$$A:\delta > 1.$$

We will consider some special alternatives and obtain the size and the relative efficiency of the Test $\hat{T}$ with respect to the corresponding best test for each of these alternatives. These results are presented in Table 1.

TABLE 1

Population	Size of $\hat{T}$ test	Relative efficiency of $\hat{T}$ test w. r. t. the corresponding best test for—		
		$\delta = 2$	$\delta = 3$	$\delta = 4$
Normal	0.23	0.83	0.76	
Uniform	0.25	0.70	0.68	0.68
Double exponential	0.25	0.92	0.81	0.81

From the above results we see that the size of the test remains more or less constant. The test is highly efficient for exponential alternatives and moderately so for normal and uniform alternatives.

7. Acknowledgment. I wish to express my deepest gratitude to Professors Erich Lehmann and Lucien LeCum, who encouraged me to work on this problem, suggested the topic and gave generous help and guidance during the course of the entire work.

REFERENCES

[1] D. A. S. FRASER, "Non-parametric theory: scale and location parameters," *Can. J. Math.*, Vol. 6 (1954), pp. 45–68.

[2] Balkrishna V. Sukhatme, "On certain two sample nonparametric tests for comparing variances," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 189–194.

[3] A. Mood, "On the asymptotic efficiency of certain nonparametric two-sample tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.

[4] E. S. Pearson, "The analysis of variance in cases of non-normal variation," *Biometrika*, Vol. 23 (1931), pp. 114–134.

[5] R. C. Geary, "Testing for normality," *Biometrika*, Vol. 34 (1947), pp. 209–241.

[6] A. K. Gayen, "The distribution of the variance ratio in random samples of any size drawn from non-normal universes," *Biometrika*, Vol. 37 (1950), pp. 236–255.

[7] G. E. P. Box, "Non-normality and tests on variances," *Biometrika*, Vol. 40 (1953), pp. 318–335.

[8] G. E. P. Box and S. L. Anderson, "Robust tests for variances and effect on non-normality and variance heterogeneity on standard tests," Institute of Statistics, North Carolina, Mimeographed Series No. 101 (1954).

[9] E. L. Lehmann, unpublished.

[10] W. Hoeffding, "A class of statistics with asymptotically normal distribution," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293–325.

[11] Balkrishna V. Sukhatme, "On the joint asymptotic distribution of the median and a U-statistic," J. Roy. Stat. Soc., B, Vol. 19 (1957), pp. 144–148.

[12] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

ESTIMATION OF LOCATION AND SCALE PARAMETERS BY ORDER STATISTICS FROM SINGLY AND DOUBLY CENSORED SAMPLES[1] PART II. TABLES FOR THE NORMAL DISTRIBUTION FOR SAMPLES OF SIZE $11 \leqq n \leqq 15$

By A. E. Sarhan and B. G. Greenberg

University of North Carolina

1. Introduction. In a previous paper [2], estimation of the mean and standard deviation from singly and doubly censored samples drawn from the normal distribution were considered for samples $n \leqq 10$. The generalization of an alternative estimate for these parameters was also obtained.

In the present work, all calculations and tables obtained for the corresponding items in Part I are extended up to $n \leqq 15$.

The method is to obtain the best linear unbiased estimates of the mean and standard deviation by taking the best linear combination of the ordered observations. The variances and covariances of the order statistics for samples $11 \leqq n \leqq 15$ which are required in carrying out these calculations are obtained from Table I in [2].

Further investigation of the efficiency of the alternative estimate under varied degrees of censoring shows that the alternative estimate proposed by Gupta [1] is better than previously supposed when judged by doubly censored samples rather than singly censored samples alone.

2. Tables. Table I gives the coefficients for the best linear estimates of the mean and standard deviation for the normal population from samples of size $11 \leqq n \leqq 15$ undergoing all possible conditions of Type II censoring. Estimation from complete or singly censored samples are simply special cases and are given in the table

$$(r_1 = r_2 = 0, \text{and } r_1 \text{ or } r_2 = 0).$$

The best linear estimates of the mean and standard deviations are obtained by using

$$\mu^* = \sum_{i=r_1+1}^{n-r_2} a_{1i} y_{(i)} ,$$

$$\sigma^* = \sum_{i=r_1+1}^{n-r_2} a_{2i} y_{(i)} ,$$

where

$$y_{(1)} < y_{(2)} < \cdots , < y_{(n)} .$$

Received August 29, 1956; revised September 19, 1957.

[1] Sponsored by the Office of Ordnance Research, U. S. Army.

n = 11

Table I. The coefficients of the most efficient linear systematic statistics of the mean and standard deviation in censored samples of sizes 11 ≤ n ≤ 15 from normal populations

$r_1 = 0$, r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$		$r_1 = 1$, r_2
0	μ^*	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909		
	σ^*	-.1883	-.1115	-.0760	-.0481	-.0234	.0000	.0234	.0481	.0760	.1115	.1883		
1	μ^*	.0781	.0841	.0869	.0891	.0910	.0928	.0945	.0963	.0982	.1891			
	σ^*	-.2149	-.1256	-.0843	-.0519	-.0233	.0038	.0309	.0593	.0911	.3149			
2	μ^*	.0592	.0744	.0814	.0869	.0917	.0962	.1005	.1049	.3047	3.1882	-2.1882	μ^*	8
	σ^*	-.2463	-.1417	-.0934	-.0555	-.0220	.0095	.0409	.0736	.4349	3.0023	-3.0023	σ^*	
3	μ^*	.0320	.0609	.0741	.0845	.0935	.1020	.1101	.4430	1.8131	-.0776	-.7355	μ^*	7
	σ^*	-.2852	-.1610	-.1038	-.0589	-.0194	.0178	.0545	.5562	1.8473	-.3251	-1.5222	σ^*	
4	μ^*	-.0082	.0415	.0642	.0820	.0974	.1116	.6116	1.2227	.0631	.0019	-.2877	μ^*	6
	σ^*	-.3357	-.1854	-.1163	-.0621	-.0146	.0299	.6842	1.3689	-.1119	-.2362	-1.0208	σ^*	
5	μ^*	-.0698	.0128	.0504	.0797	.1049	.8220	.8646	.1026	.0738	.0407	-.0818	μ^*	5
	σ^*	-.4045	-.2175	-.1317	-.0647	-.0061	.8246	1.0717	-.0195	-.0985	-.1881	-.7656	σ^*	
6	μ^*	-.1702	-.0323	.0303	.0786	1.0937	.6157	.1116	.0971	.0813	.0632	.0311	μ^*	4
	σ^*	-.5053	-.2627	-.1519	-.0657	.9857	.8517	.0301	-.0272	-.0882	-.1569	-.6095	σ^*	
7	μ^*	-.3516	-.1104	-.0016	1.4636	.4301	.1093	.1022	.0948	.0868	.0776	.0992	μ^*	3
	σ^*	-.6687	-.3331	-.1807	1.1825	.6715	.0611	.0154	-.0307	-.0796	-.1345	-.5032	σ^*	
8	μ^*	-.7445	-.2712	2.0157	.2859	.1025	.0998	.0970	.0941	.0909	.0872	.1426	μ^*	2
	σ^*	-.9862	-.4630	1.4492	.5130	.0836	.0410	.0061	-.0320	-.0722	-.1173	-.4252	σ^*	
9	μ^*	-2.0245	3.0245	.1710	.0938	.0940	.0941	.0942	.0941	.0940	.0938	.1710	μ^*	1
	σ^*	-1.9065	1.9065	.3647	.1033	.0656	.0320	.0000	-.0320	-.0656	-.1033	-.3647	σ^*	
				$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$		

$r_1 = 2$, r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$		$r_1 = 3$, r_2
2	μ^*	.2529	.0986	.0990	.0990	.0990	.0986	.2529		
	σ^*	-.6113	-.0952	-.0465	.0000	.0465	.0952	.6113		
3	μ^*	.1978	.0927	.0983	.1032	.1079	.4001			
	σ^*	-.7445	-.1095	-.0482	.0102	.0682	.8239			
4	μ^*	.1031	.0834	.0981	.1117	.6036	1.9486	-.9486	μ^*	6
	σ^*	-.9395	-.1287	-.0484	.0276	1.0890	4.2179	-4.2179	σ^*	
5	μ^*	-.0736	.0675	.0999	.9062	.9504	.0966	-.0470	μ^*	5
	σ^*	-1.2584	-.1573	-.0453	1.4610	2.2169	-.1019	-2.1150	σ^*	
6	μ^*	-.4631	.0359	1.4273	.5616	.1125	.1024	.2235	μ^*	4
	σ^*	-1.8863	-.2083	2.0946	1.4723	.0202	-.0867	-1.4057	σ^*	
7	μ^*	-1.7312	2.7312	.3413	.1057	.1059	.1057	.3413	μ^*	3
	σ^*	-3.7473	3.7473	1.0451	.0764	.0000	-.0764	-1.0451	σ^*	
				$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$		

$r_1 = 4$, r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$
4	μ^*	.4421	.1157	.4421
	σ^*	-2.2233	.0000	2.2233
5	μ^*	.0000	1.0000	
	σ^*	-4.4466	4.4466	

n = 12

Table I. (Continued)

$r_1 = 0$, r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$		$r_1 = 1$, r_2
0	μ^*	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833		
	σ^*	-.1748	-.1061	-.0749	-.0506	-.0294	-.0097	.0097	.0294	.0506	.0749	.1061	.1748		
1	μ^*	.0726	.0775	.0796	.0813	.0828	.0842	.0855	.0868	.0882	.0896	.1719			
	σ^*	-.1972	-.1185	-.0827	-.0548	-.0305	-.0079	.0142	.0367	.0608	.0881	.2919			
2	μ^*	.0574	.0693	.0747	.0789	.0825	.0859	.0891	.0923	.0956	.2745	3.4554	-2.1554	μ^*	9
	σ^*	-.2232	-.1324	-.0911	-.0590	-.0310	-.0050	.0203	.0461	.0733	.4020	3.0970	-3.0970	σ^*	
3	μ^*	.0360	.0581	.0682	.0759	.0827	.0888	.0948	.1006	.3950	1.9966	-.1242	-.8725	μ^*	8
	σ^*	-.2545	-.1487	-.1007	-.0633	-.0308	-.0007	.0286	.0582	.5119	1.9252	-.3546	-1.5706	σ^*	
4	μ^*	.0057	.0428	.0595	.0724	.0836	.0938	.1036	.5386	1.3737	.0373	-.0298	-.3812	μ^*	7
	σ^*	-.2937	-.1686	-.1119	-.0678	-.0296	.0058	.0400	.6259	1.4442	-.1345	-.2554	-1.0542	σ^*	
5	μ^*	-.0382	.0210	.0477	.0684	.0861	.1022	.7128	.9961	.0870	.0542	.0162	-.1536	μ^*	6
	σ^*	-.3448	-.1939	-.1255	-.0726	-.0267	.0155	.7479	1.1482	-.0390	-.1151	-.2024	-.7918	σ^*	
6	μ^*	-.1048	-.0109	.0313	.0637	.0915	.9292	.7331	.1018	.0844	.0652	.0430	-.0275	μ^*	5
	σ^*	-.4146	-.2274	-.1428	-.0774	-.0210	.8833	.9313	.0121	-.0423	-.1011	-.1684	-.6316	σ^*	
7	μ^*	-.2125	-.0609	.0070	.0589	1.2075	.5361	.1035	.0939	.0839	.0729	.0601	.0496	μ^*	4
	σ^*	-.5171	-.2749	-.1659	-.0820	1.0399	.7559	.0435	.0009	-.0430	-.0903	-.1442	-.5228	σ^*	
8	μ^*	-.4059	-.1472	-.0321	1.5852	.3819	.0996	.0916	.0896	.0843	.0785	.0717	.0998	μ^*	3
	σ^*	-.6836	-.3493	-.1996	1.2324	.6042	.0654	.0295	-.0059	-.0423	-.0814	-.1259	-.4435	σ^*	
9	μ^*	-.8225	-.3249	2.1474	.2580	.0932	.0912	.0892	.0872	.0851	.0827	.0798	.1336	μ^*	2
	σ^*	-1.0075	-.4874	1.4948	.4660	.0827	.0503	.0198	-.0102	-.0410	-.0738	-.1112	-.3825	σ^*	
10	μ^*	-2.1728	3.1728	.1566	.0856	.0858	.0859	.0860	.0860	.0859	.0858	.0856	.1566	μ^*	1
	σ^*	-1.9474	1.9474	.3334	.0989	.0671	.0391	.0128	-.0128	-.0391	-.0671	-.0989	-.3334	σ^*	
				$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$		

$r_1 = 2$, r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$		$r_1 = 3$, r_2
2	μ^*	.2309	.0895	.0898	.0899	.0899	.0898	.0895	.2309		
	σ^*	-.5437	-.0933	-.0545	-.0179	.0179	.0545	.0933	.5437		
3	μ^*	.1883	.0840	.0880	.0914	.0947	.0978	.3558			
	σ^*	-.6440	-.1061	-.0587	-.0142	.0292	.0733	.7204			
4	μ^*	.1206	.0759	.0858	.0947	.1032	.5197	2.3903	-1.3903	μ^*	7
	σ^*	-.7811	-.1225	-.0631	-.0076	.0465	.9278	4.4525	-4.4525	σ^*	
5	μ^*	.0076	.0633	.0832	.1015	.7444	1.2001	.0699	-.2700	μ^*	6
	σ^*	-.9829	-.1451	-.0676	.0045	1.1911	2.3763	-.1421	-2.2342	σ^*	
6	μ^*	-.1993	.0417	.0806	1.0771	.7412	.1016	.0836	.0736	μ^*	5
	σ^*	-1.3139	-.1796	-.0719	1.5654	1.6109	-.0111	-.1124	-1.4874	σ^*	
7	μ^*	-.6469	-.0018	1.6507	.4812	.1029	.0970	.0907	.2282	μ^*	4
	σ^*	-1.9673	-.2129	2.2102	1.1775	.0488	-.0223	-.0950	-1.1090	σ^*	
8	μ^*	-2.0971	3.0971	.3097	.0950	.0953	.0953	.0950	.3097	μ^*	3
	σ^*	-3.9063	3.9063	.8780	.0828	.0273	-.0273	-.0828	-.8780	σ^*	
				$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$		

$r_1 = 4$, r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$		$r_1 = 5$, r_2
4	μ^*	.3973	.1027	.1027	.3973		
	σ^*	-1.5849	-.0497	.0497	1.5849		
5	μ^*	.1962	.1034	.7004			
	σ^*	-2.3874	-.0468	2.4342			
6	μ^*	-.4893	1.4893	.5000	.5000	μ^*	5
	σ^*	-4.7696	4.7696	4.8738	-4.8738	σ^*	
				$y_{(7)}$	$y_{(6)}$		

$n = 13$

Table I. (Continued)

$r_1 = 0$ r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$		$r_1 = 1$ r_2
0	μ^*	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769		
	σ^*	-.1632	-.1013	-.0735	-.0520	-.0335	-.0164	.0000	.0164	.0335	.0520	.0735	.1013	.1632		
1	μ^*	.0679	.0718	.0735	.0749	.0761	.0771	.0781	.0792	.0802	.0813	.0824	.1576			
	σ^*	-.1824	-.1122	-.0806	-.0563	-.0353	-.0160	.0026	.0212	.0404	.0612	.0850	.2724			
2	μ^*	.0552	.0648	.0691	.0724	.0752	.0778	.0803	.0827	.0852	.0877	.2497	3.7044	-2.7044	μ^*	10
	σ^*	-.2043	-.1243	-.0884	-.0607	-.0368	-.0148	.0063	.0273	.0490	.0723	.3743	3.1823	-3.1823	σ^*	
3	μ^*	.0380	.0555	.0633	.0693	.0745	.0792	.0836	.0880	.0924	.3564	2.1686	-.1687	-.9999	μ^*	9
	σ^*	-.2301	-.1382	-.0970	-.0653	-.0379	-.0128	.0113	.0352	.0598	.4750	1.9951	-.3810	-1.6141	σ^*	
4	μ^*	.0144	.0430	.0557	.0655	.0739	.0816	.0888	.0958	.4813	1.5160	.0123	-.0600	-.4683	μ^*	8
	σ^*	-.2616	-.1549	-.1071	-.0703	-.0386	-.0095	.0182	.0456	.5781	1.5114	-.1546	-.2727	-1.0841	σ^*	
5	μ^*	-.0185	.0259	.0457	.0610	.0740	.0857	.0968	.6294	1.1209	.0712	.0352	-.0070	-.2203	μ^*	7
	σ^*	-.3011	-.1754	-.1191	-.0758	-.0386	-.0046	.0278	.6867	1.2161	-.0561	-.1298	-.2151	-.8151	σ^*	
6	μ^*	-.0659	.0020	.0322	.0553	.0750	.0928	.8085	.8453	.0916	.0716	.0496	.0239	-.0820	μ^*	6
	σ^*	-.3528	-.2015	-.1339	-.0819	-.0374	.0032	.8042	1.0015	-.0034	-.0556	-.1127	-.1786	-.6512	σ^*	
7	μ^*	-.1371	-.0330	.0132	.0484	.0784	1.0301	.6382	.0969	.0853	.0730	.0595	.0437	.0035	μ^*	5
	σ^*	-.4236	-.2363	-.1528	-.0888	-.0341	.9355	.8296	.0287	-.0116	-.0538	-.0998	-.1528	-.5401	σ^*	
8	μ^*	-.2516	-.0876	-.0151	.0400	1.3143	.4755	.0955	.0889	.0820	.0746	.0666	.0572	.0597	μ^*	4
	σ^*	-.5276	-.2859	-.1785	-.0964	1.0884	.6826	.0505	.0172	-.0164	-.0515	-.0896	-.1335	-.4594	σ^*	
9	μ^*	-.4561	-.1817	-.0610	1.6988	.3439	.0912	.0876	.0840	.0802	.0763	.0719	.0668	.0983	μ^*	3
	σ^*	-.6969	-.3638	-.2165	1.2773	.5509	.0669	.0377	.0093	-.0192	-.0489	-.0811	-.1181	-.3976	σ^*	
10	μ^*	-.8946	-.3753	2.2699	.2353	.0853	.0839	.0824	.0809	.0794	.0778	.0759	.0737	.1254	μ^*	2
	σ^*	-1.0266	-.5094	1.5360	.4278	.0808	.0536	.0283	.0039	-.0207	-.0462	-.0738	-.1055	-.3483	σ^*	
11	μ^*	-2.3101	3.3101	.1445	.0788	.0790	.0791	.0791	.0791	.0791	.0791	.0790	.0788	.1445	μ^*	1
	σ^*	-1.9845	1.9845	.3076	.0947	.0673	.0434	.0213	.0000	-.0213	-.0434	-.0673	-.0947	-.3076	σ^*	
				$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$		

n = 13 Table I. (Continued)

$r_1 = 2$ r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$		$r_1 = 3$ r_2
2	μ^*	.2125	.0819	.0822	.0823	.0823	.0823	.0822	.0819	.2125		
	σ^*	-.4911	-.0904	-.0584	-.0287	.0000	.0287	.0584	.0904	.4911		
3	μ^*	.1784	.0770	.0799	.0825	.0848	.0871	.0892	.3211			
	σ^*	-.5698	-.1017	-.0636	-.0283	.0058	.0398	.0748	.6431			
4	μ^*	.1272	.0700	.0770	.0833	.0892	.0949	.4584	2.8102	-1.8102	μ^*	8
	σ^*	-.6724	-.1157	-.0695	-.0266	.0146	.0555	.8142	4.6615	-4.6615	σ^*	
5	μ^*	.0479	.0597	.0732	.0854	.0969	.6370	1.4405	.0415	-.4820	μ^*	7
	σ^*	-.8133	-.1341	-.0763	-.0229	.0282	1.0184	2.5175	-.1774	-2.3402	σ^*	
6	μ^*	-.0822	.0437	.0680	.0901	.8805	.9169	.0883	.0639	-.0691	μ^*	6
	σ^*	-1.0214	-.1597	-.0846	-.0154	1.2810	1.7328	-.0380	-.1350	-1.5598	σ^*	
7	μ^*	-.3174	.0164	.0607	1.2404	.6209	.0970	.0865	.0752	.1204	μ^*	5
	σ^*	-1.3635	-.1994	-.0952	1.6581	1.2929	.0255	-.0417	-.1115	-1.1652	σ^*	
8	μ^*	-.8233	-.0391	1.8624	.4246	.0939	.0902	.0862	.0820	.2231	μ^*	4
	σ^*	-2.0397	-.2735	2.3132	.9927	.0610	.0095	-.0423	-.0958	-.9251	σ^*	
9	μ^*	-2.4408	3.4408	.2838	.0863	.0866	.0866	.0866	.0863	.2838	μ^*	3
	σ^*	-4.0488	4.0488	.7622	.0841	.0414	.0000	-.0414	-.0841	-.7622	σ^*	
				$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$		

$r_1 = 4$ r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$		$r_1 = 5$ r_2
4	μ^*	.3615	.0923	.0924	.0923	.3615		
	σ^*	-1.2545	-.0674	.0000	.0674	1.2545		
5	μ^*	.2377	.0900	.0977	.5745			
	σ^*	-1.6839	-.0755	.0184	1.7410			
6	μ^*	-.0422	.0861	.9562	1.0000	.0000	μ^*	6
	σ^*	-2.5323	-.0873	2.6196	5.2487	-5.2487	σ^*	
7	μ^*	-.9632	1.9632	.4498	.1003	.4498	μ^*	5
	σ^*	-5.0555	5.0555	2.6243	.0000	-2.6243	σ^*	
				$y_{(8)}$	$y_{(7)}$	$y_{(6)}$		

n = 14

Table I. (Continued)

$r_1 = 0$, r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$	$y_{(14)}$		$r_1 = 1$, r_2
0	μ^*	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714		
	σ^*	-.1532	-.0968	-.0717	-.0526	-.0362	-.0212	-.0070	.0070	.0212	.0362	.0526	.0717	.0968	.1532		
1	μ^*	.0637	.0669	.0683	.0694	.0704	.0712	.0721	.0728	.0736	.0745	.0753	.0762	.1455			
	σ^*	-.1698	-.1065	-.0784	-.0568	-.0384	-.0216	-.0056	.0100	.0259	.0426	.0609	.0820	.2556			
2	μ^*	.0530	.0609	.0643	.0670	.0692	.0713	.0732	.0751	.0770	.0789	.0809	.2291	3.9374	-2.9374	μ^*	11
	σ^*	-.1885	-.1171	-.0854	-.0612	-.0404	-.0215	-.0036	.0140	.0319	.0505	.0707	.3506	3.2597	-3.2597	σ^*	
3	μ^*	.0388	.0529	.0592	.0639	.0680	.0717	.0752	.0785	.0819	.0852	.3247	2.3304	-.2112	-1.1191	μ^*	10
	σ^*	-.2102	-.1292	-.0933	-.0658	-.0423	-.0209	-.0006	.0192	.0393	.0601	.4438	2.0584	-.4048	-1.6536	σ^*	
4	μ^*	.0199	.0426	.0526	.0602	.0667	.0726	.0782	.0835	.0887	.4350	1.6505	-.0121	-.0888	-.5496	μ^*	9
	σ^*	-.2361	-.1434	-.1023	-.0709	-.0440	-.0196	.0035	.0260	.0487	.5382	1.5721	-.1726	-.2882	-1.1112	σ^*	
5	μ^*	-.0057	.0288	.0440	.0557	.0655	.0744	.0828	.0908	.5637	1.2394	.0556	.0167	-.0291	-.2827	μ^*	8
	σ^*	-.2678	-.1604	-.1129	-.0765	-.0455	-.0174	.0092	.0350	.6363	1.2771	-.0713	-.1430	-.2267	-.8362	σ^*	
6	μ^*	-.0411	.0102	.0328	.0500	.0646	.0777	.0899	.7159	.9525	.0811	.0591	.0345	.0057	-.1329	μ^*	7
	σ^*	-.3077	-.1815	-.1256	-.0829	-.0466	-.0137	.0172	.7407	1.0642	-.0171	-.0675	-.1231	-.1878	-.6688	σ^*	
7	μ^*	-.0915	-.0158	.0175	.0429	.0643	.0835	.8992	.7365	.0897	.0765	.0623	.0466	.0281	-.0397	μ^*	6
	σ^*	-.3599	-.2084	-.1414	-.0903	-.0469	-.0077	.8546	.8949	.0158	-.0227	-.0635	-.1084	-.1606	-.5555	σ^*	
8	μ^*	-.1670	-.0537	-.0040	.0338	.0655	1.1255	.5661	.0908	.0827	.0743	.0652	.0552	.0435	.0222	μ^*	5
	σ^*	-.4317	-.2444	-.1618	-.0990	-.0457	.9825	.7515	.0379	.0065	-.0256	-.0596	-.0970	-.1402	-.4734	σ^*	
9	μ^*	-.2879	-.1127	-.0360	.0218	1.4148	.4277	.0883	.0834	.0784	.0733	.0677	.0616	.0544	.0651	μ^*	4
	σ^*	-.5372	-.2959	-.1898	-.1094	1.1322	.6244	.0540	.0270	.0003	-.0271	-.0560	-.0876	-.1242	-.4107	σ^*	
10	μ^*	-.5027	-.2142	-.0886	1.8054	.3129	.0840	.0812	.0785	.0758	.0729	.0699	.0665	.0625	.0957	μ^*	3
	σ^*	-.7091	-.3771	-.2318	1.3180	.5075	.0669	.0426	.0191	-.0040	-.0276	-.0525	-.0797	-.1112	-.3610	σ^*	
11	μ^*	-.9616	-.4228	2.3843	.2163	.0787	.0776	.0765	.0753	.0742	.0730	.0717	.0703	.0685	.1180	μ^*	2
	σ^*	-1.0441	-.5293	1.5734	.3961	.0795	.0552	.0338	.0133	-.0070	-.0275	-.0492	-.0728	-.1002	-.3203	σ^*	
12	μ^*	-2.4378	3.4378	.1341	.0730	.0731	.0732	.0733	.0733	.0733	.0733	.0732	.0731	.0730	.1341	μ^*	1
	σ^*	-2.0182	2.0182	.2858	.0907	.0667	.0459	.0270	.0089	-.0089	-.0270	-.0459	-.0667	-.0907	-.2858	σ^*	
				$y_{(13)}$	$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$		

n = 14

Table I. (Continued)

$r_1 = 2$ r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$		$r_1 = 3$ r_2
2	μ^*	.1969	.0755	.0758	.0759	.0759	.0759	.0759	.0758	.0755	.1969		
	σ^*	-.4488	-.0872	-.0602	-.0354	-.0117	.0117	.0354	.0602	.0872	.4488		
3	μ^*	.1689	.0711	.0734	.0753	.0771	.0788	.0804	.0820	.2930			
	σ^*	-.5125	-.0972	-.0656	-.0367	-.0090	.0182	.0457	.0744	.5827			
4	μ^*	.1287	.0651	.0703	.0749	.0792	.0834	.0874	.4111	3.2094	-2.2094	μ^*	9
	σ^*	-.5926	-.1092	-.0718	-.0375	-.0048	.0273	.0596	.7291	4.8497	-4.8497	σ^*	
5	μ^*	.0695	.0565	.0661	.0748	.0829	.0907	.5593	1.6712	.0123	-.6835	μ^*	8
	σ^*	-.6975	-.1244	-.0791	-.0376	.0018	.0403	.8965	2.6442	-.2087	-2.4355	σ^*	
6	μ^*	-.0206	.0440	.0606	.0755	.0895	.7510	1.0874	.0734	.0439	-.2047	μ^*	7
	σ^*	-.8420	-.1444	-.0881	-.0365	.0123	1.0987	1.8416	-.0617	-.1552	-1.6247	σ^*	
7	μ^*	-.1668	.0247	.0527	.0779	1.0115	.7583	.0893	.0750	.0596	.0178	μ^*	6
	σ^*	-1.0559	-.1728	-.0996	-.0330	1.3612	1.3951	.0053	-.0588	-.1262	-1.2153	σ^*	
8	μ^*	-.4285	-.0084	.0405	1.3964	.5397	.0905	.0837	.0766	.0690	.1405	μ^*	5
	σ^*	-1.4081	-.2172	-.1160	1.7413	1.0932	.0424	-.0061	-.0555	-.1074	-.9667	σ^*	
9	μ^*	-.9875	-.0756	2.0631	.3818	.0860	.0835	.0809	.0781	.0750	.2147	μ^*	4
	σ^*	-2.1051	-.3010	2.4061	.8643	.0660	.0266	-.0125	-.0522	-.0937	-.7987	σ^*	
10	μ^*	-2.7647	3.7647	.2621	.0791	.0794	.0795	.0795	.0794	.0791	.2621	μ^*	3
	σ^*	-4.1778	4.1778	.6765	.0831	.0489	.0162	-.0162	-.0489	-.0831	-.6765	σ^*	
				$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$		

$r_1 = 4$ r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$		$r_1 = 5$ r_2
4	μ^*	.3320	.0839	.0841	.0841	.0839	.3320		
	σ^*	-1.0495	-.0738	-.0244	.0244	.0738	1.0495		
5	μ^*	.2456	.0807	.0857	.0903	.4978			
	σ^*	-1.3227	-.0839	-.0195	.0439	1.3821			
6	μ^*	.0837	.0755	.0895	.7512	1.4921	-.4921	μ^*	7
	σ^*	-1.7721	-.0983	-.0088	1.8791	5.5823	-5.5823	σ^*	
7	μ^*	-.2720	.0657	1.2063	.7061	.0910	.2029	μ^*	6
	σ^*	-2.6618	-.1229	2.7848	2.8338	-.0406	-2.7931	σ^*	
8	μ^*	-1.4198	2.4198	.4098	.0902	.0902	.4098	μ^*	5
	σ^*	-5.3115	5.3115	1.8560	.0441	-.0441	-1.8560	σ^*	
				$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$		

$r_1 = 6$ r_2		$y_{(7)}$	$y_{(8)}$
6	μ^*	.5000	.5000
	σ^*	-5.6716	5.6716

n = 15

Table I. (Continued)

$r_1 = 0$, r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$	$y_{(14)}$	$y_{(15)}$		$r_1 = 1$, r
0	μ^*	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667		
	σ^*	-.1444	-.0927	-.0699	-.0526	-.0379	-.0247	-.0122	.0000	.0122	.0247	.0379	.0526	.0699	.0927	.1444		
1	μ^*	.0599	.0627	.0639	.0648	.0655	.0662	.0669	.0675	.0682	.0688	.0695	.0702	.0709	.1351			
	σ^*	-.1590	-.1013	-.0760	-.0568	-.0404	-.0256	-.0116	.0019	.0154	.0293	.0440	.0602	.0791	.2409			
2	μ^*	.0508	.0574	.0602	.0624	.0642	.0659	.0675	.0690	.0704	.0719	.0735	.0751	.2116	4.1564	-3.1564	μ^*	12
	σ^*	-.1752	-.1108	-.0825	-.0610	-.0427	-.0262	-.0106	.0044	.0195	.0349	.0512	.0690	.3300	3.3306	-3.3306	σ^*	
3	μ^*	.0390	.0506	.0556	.0595	.0628	.0657	.0685	.0711	.0737	.0763	.0790	.2982	2.4829	-.2519	-1.2311	μ^*	11
	σ^*	-.1937	-.1214	-.0897	-.0655	-.0450	-.0265	-.0091	.0078	.0246	.0417	.0598	.4169	2.1162	-.4265	-1.6897	σ^*	
4	μ^*	.0234	.0418	.0498	.0560	.0611	.0658	.0701	.0743	.0784	.0824	.3969	1.7779	-.0356	-.1163	-.6260	μ^*	10
	σ^*	-.2154	-.1336	-.0977	-.0705	-.0473	-.0264	-.0068	.0122	.0310	.0502	.5042	1.6273	-.1890	-.3024	-1.1359	σ^*	
5	μ^*	.0030	.0305	.0425	.0516	.0593	.0663	.0727	.0789	.0849	.5104	1.3522	.0403	-.0011	-.0501	-.3413	μ^*	9
	σ^*	-.2414	-.1481	-.1071	-.0760	-.0496	-.0258	-.0035	.0180	.0393	.5940	1.3324	-.0850	-.1549	-.2373	-.8553	σ^*	
6	μ^*	-.0244	.0155	.0330	.0462	.0574	.0674	.0767	.0856	.6425	1.0549	.0706	.0467	.0200	-.0115	-.1807	μ^*	8
	σ^*	-.2733	-.1654	-.1181	-.0822	-.0518	-.0244	.0012	.0258	.6882	1.1208	-.0293	-.0781	-.1325	-.1963	-.6847	σ^*	
7	μ^*	-.0621	-.0046	.0205	.0395	.0555	.0698	.0830	.7983	.8308	.0823	.0677	.0519	.0342	.0134	-.0803	μ^*	7
	σ^*	-.3136	-.1870	-.1315	-.0894	-.0538	-.0219	.0079	.7892	.9536	.0045	-.0326	-.0722	-.1162	-.1677	-.5694	σ^*	
8	μ^*	-.1155	-.0326	.0036	.0309	.0539	.0743	.9854	.6537	.0857	.0764	.0666	.0561	.0443	.0305	-.0131	μ^*	6
	σ^*	-.3664	-.2146	-.1482	-.0979	-.0555	-.0174	.9001	.8129	.0269	-.0029	-.0339	-.0670	-.1036	-.1464	-.4859	σ^*	
9	μ^*	-.1950	-.0732	-.0203	.0196	.0531	1.2157	.5093	.0848	.0789	.0728	.0664	.0595	.0518	.0428	.0338	μ^*	5
	σ^*	-.4390	-.2518	-.1700	-.1082	-.0562	1.0252	.6892	.0430	.0177	-.0078	-.0342	-.0624	-.0935	-.1298	-.4223	σ^*	
10	μ^*	-.3217	-.1364	-.0560	.0043	1.5097	.3890	.0818	.0781	.0744	.0706	.0666	.0623	.0575	.0519	.0677	μ^*	4
	σ^*	-.5459	-.3050	-.2002	-.1211	1.1722	.5767	.0555	.0331	.0111	-.0110	-.0339	-.0582	-.0850	-.1163	-.3721	σ^*	
11	μ^*	-.5462	-.2448	-.1148	1.9058	.2873	.0777	.0756	.0735	.0714	.0693	.0671	.0647	.0620	.0587	.0927	μ^*	3
	σ^*	-.7201	-.3892	-.2458	1.3552	.4712	.0661	.0453	.0256	.0062	-.0131	-.0331	-.0543	-.0778	-.1050	-.3311	σ^*	
12	μ^*	-1.0242	-.4676	2.4918	.2002	.0730	.0721	.0712	.0704	.0695	.0686	.0676	.0666	.0654	.0640	.1113	μ^*	2
	σ^*	-1.0601	-.5477	1.6077	.3693	.0761	.0558	.0373	.0197	.0026	-.0145	-.0321	-.0508	-.0714	-.0954	-.2968	σ^*	
13	μ^*	-2.5574	3.5574	.1252	.0679	.0681	.0682	.0682	.0682	.0682	.0682	.0682	.0682	.0681	.0679	.1252	μ^*	1
	σ^*	-2.0493	2.0493	.2672	.0869	.0656	.0474	.0308	.0152	.0000	-.0152	-.0308	-.0474	-.0656	-.0869	-.2672	σ^*	
				$y_{(14)}$	$y_{(13)}$	$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$		

Table I. (Continued)

n = 15

$r_1 = 2$, r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$		$r_1 = 3$, r_2
2	μ^*	.1835	.0701	.0703	.0704	.0705	.0705	.0705	.0704	.0703	.0701	.1835		
	σ^*	-.4139	-.0838	-.0607	-.0395	-.0195	.0000	.0195	.0395	.0607	.0838	.4139		
3	μ^*	.1601	.0662	.0679	.0695	.0708	.0721	.0734	.0746	.0758	.2697			
	σ^*	-.4668	-.0927	-.0660	-.0417	-.0187	.0037	.0261	.0490	.0730	.5341			
4	μ^*	.1274	.0609	.0649	.0684	.0717	.0748	.0778	.0808	.3734	3.5892	-2.5892	μ^*	10
	σ^*	-.5314	-.1032	-.0721	-.0438	-.0170	.0090	.0349	.0612	.6624	5.0207	-5.0207	σ^*	
5	μ^*	.0815	.0536	.0608	.0673	.0733	.0790	.0845	.5000	1.8923	-.0171	-.8752	μ^*	9
	σ^*	-.6129	-.1160	-.0792	-.0457	-.0112	.0164	.0468	.8049	2.7587	-.2368	-2.5219	σ^*	
6	μ^*	.0150	.0436	.0555	.0662	.0762	.0857	.6579	1.2523	.0577	.0238	-.3338	μ^*	8
	σ^*	-.7200	-.1323	-.0878	-.0475	-.0095	.0272	.9699	1.9396	-.0828	-.1734	-1.6834	σ^*	
7	μ^*	-.0854	.0289	.0481	.0654	.0815	.8614	.8925	.0804	.0629	.0441	-.0799	μ^*	7
	σ^*	-.8679	-.1539	-.0987	-.0486	-.0017	1.1708	1.1867	-.0125	-.0741	-.1395	-1.2605	σ^*	
8	μ^*	-.2467	.0063	.0375	.0655	1.1375	.6535	.0857	.0764	.0666	.0561	.0617	μ^*	6
	σ^*	-1.0872	-.1847	-.1132	-.0486	1.1336	1.1828	.0264	-.0198	-.0674	-.1178	-1.0042	σ^*	
9	μ^*	-.5335	-.0324	.0205	1.5454	.4801	.0840	.0793	.0745	.0694	.0639	.1487	μ^*	5
	σ^*	-1.4486	-.2334	-.1346	1.8166	.9544	.0508	.0137	-.0235	-.0619	-.1023	-.8312	σ^*	
10	μ^*	-1.1426	-.1111	2.2537	.3478	.0792	.0775	.0756	.0737	.0716	.0693	.2054	μ^*	4
	σ^*	-2.1646	-.3259	2.4905	.7692	.0678	.0363	.0055	-.0254	-.0571	-.0905	-.7058	σ^*	
11	μ^*	-3.0706	4.0706	.2436	.0731	.0733	.0734	.0734	.0734	.0733	.0731	.2436	μ^*	3
	σ^*	-4.2953	4.2953	.6102	.0810	.0528	.0261	.0000	-.0261	-.0528	-.0810	-.6102	σ^*	
				$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$		

$r_1 = 4$, r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$		$r_1 = 5$, r_2
4	μ^*	.3073	.0770	.0772	.0772	.0772	.0770	.3073		
	σ^*	-.9085	-.0756	-.0374	.0000	.0374	.0756	.9085		
5	μ^*	.2424	.0736	.0770	.0802	.0832	.4435			
	σ^*	-1.1001	-.0854	-.0377	.0088	.0552	1.1592			
6	μ^*	.1339	.0686	.0774	.0859	.6342	1.9724	-.9724	μ^*	8
	σ^*	-1.3838	-.0987	-.0367	.0236	1.4956	5.8824	-5.8824	σ^*	
7	μ^*	-.0644	.0601	.0791	.9252	.9607	.0776	-.0383	μ^*	7
	σ^*	-1.8515	-.1186	-.0327	2.0028	3.0212	-.0764	-2.9448	σ^*	
8	μ^*	-.4928	.0435	1.4494	.5846	.0865	.0804	.2486	μ^*	6
	σ^*	-2.7788	-.1547	2.9335	2.0090	.0169	-.0670	-1.9590	σ^*	
9	μ^*	-1.8586	2.8586	.3769	.0820	.0821	.0820	.3769	μ^*	5
	σ^*	-5.5431	5.5431	1.1614	.0604	.0000	-.0604	-1.1614	σ^*	
				$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$		

$r_1 = 6$, r_2		$y_{(7)}$	$y_{(8)}$	$y_{(9)}$
6	μ^*	.4557	.0887	.4557
	σ^*	-3.0248	.0000	3.0248
7	μ^*	.0000	1.0000	
	σ^*	-6.0497	6.0497	

Table II. Variances and Covariances of the Estimates of the

of Sizes $11 \leq n \leq 15$

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10
11	0	$V(\mu^*)$	.0909	.0929	.0966	.1031	.1143	.1342	.1718	.2504	.4493	1.2179	
		$V(\sigma^*)$	.0517	.0601	.0704	.0836	.1013	.1262	.1640	.2279	.3577	.7550	
		$Cov(\mu^*,\sigma^*)$	.0000	.0041	.0102	.0195	.0336	.0559	.0936	.1644	.3251	.8777	
	1	$V(\mu^*)$		.0943	.0975	.1033	.1143	.1356	.1805	.2929	.7093		
		$V(\sigma^*)$		.0712	.0854	.1043	.1311	.1717	.2408	.3817	.8140		
		$Cov(\mu^*,\sigma^*)$		.0000	.0066	.0172	.0343	.0637	.1194	.2453	.6696		
	2	$V(\mu^*)$			.0997	.1044	.1144	.1364	.1926	.3997			
		$V(\sigma^*)$			.1052	.1330	.1753	.2469	.3928	.8389			
		$Cov(\mu^*,\sigma^*)$			.0000	.0115	.0321	.0718	.1623	.4662			
	3	$V(\mu^*)$				.1075	.1153	.1369	.2202				
		$V(\sigma^*)$				.1764	.2496	.3982	.8512				
		$Cov(\mu^*,\sigma^*)$				.0000	.0239	.0806	.2748				
	4	$V(\mu^*)$					.1191	.1372					
		$V(\sigma^*)$					.3999	.8564					
		$Cov(\mu^*,\sigma^*)$					.0000	.0908					
12	0	$V(\mu^*)$	.0833	.0849	.0878	.0926	.1004	.1136	.1363	.1784	.2650	.4809	1.3044
		$V(\sigma^*)$	.0469	.0538	.0620	.0723	.0855	.1033	.1283	.1663	.2305	.3610	.7601
		$Cov(\mu^*,\sigma^*)$	.0000	.0033	.0081	.0152	.0254	.0406	.0645	.1045	.1791	.3469	.9202
	1	$V(\mu^*)$		.0861	.0885	.0929	.1004	.1139	.1393	.1915	.3195	.7864	
		$V(\sigma^*)$		.0627	.0737	.0878	.1067	.1336	.1745	.2438	.3854	.8195	
		$Cov(\mu^*,\sigma^*)$		.0000	.0052	.0130	.0250	.0440	.0762	.1364	.2710	.7212	
	2	$V(\mu^*)$			.0903	.0939	.1007	.1140	.1421	.2115	.4614		
		$V(\sigma^*)$			.0884	.1082	.1361	.1786	.2505	.3971	.8450		
		$Cov(\mu^*,\sigma^*)$			.0000	.0084	.0222	.0460	.0909	.1918	.5263		
	3	$V(\mu^*)$				.0963	.1018	.1140	.1457	.2617			
		$V(\sigma^*)$				.1369	.1804	.2539	.4032	.8581			
		$Cov(\mu^*,\sigma^*)$				.0000	.0153	.0454	.1141	.3438			
	4	$V(\mu^*)$					.1049	.1144	.1524				
		$V(\sigma^*)$					.2549	.4059	.8647				
		$Cov(\mu^*,\sigma^*)$					.0000	.0379	.1699				
	5	$V(\mu^*)$						.1175					
		$V(\sigma^*)$						.8667					
		$Cov(\mu^*,\sigma^*)$						.0000					

Mean (μ^*) and Standard Deviation (σ^*) for Censored Samples

from Normal Populations

n	r_1	r_2	0	1	2	3	4	5	6	7	8	9	10	11	12
13	0	$V(\mu^*)$	.0769	.0782	.0805	.0841	.0899	.0991	.1141	.1395	.1858	.2799	.5118	1.3868	
		$V(\sigma^*)$	.0429	.0486	.0554	.0636	.0739	.0872	.1050	.1302	.1688	.2329	.3640	.7646	
		$Cov(\mu^*, \sigma^*)$	.0000	.0027	.0066	.0121	.0198	.0309	.0472	.0725	.1146	.1925	.3668	.9589	
	1	$V(\mu^*)$		.0792	.0811	.0845	.0900	.0992	.1152	.1444	.2036	.3466	.8613		
		$V(\sigma^*)$		.0559	.0647	.0756	.0898	.1088	.1358	.1768	.2465	.3886	.8243		
		$Cov(\mu^*, \sigma^*)$		.0000	.0041	.0101	.0190	.0322	.0530	.0876	.1518	.2944	.7679		
	2	$V(\mu^*)$			.0826	.0854	.0903	.0992	.1159	.1500	.2324	.5237			
		$V(\sigma^*)$			.0761	.0909	.1106	.1387	.1813	.2536	.4007	.8502			
		$Cov(\mu^*, \sigma^*)$			.0000	.0064	.0162	.0320	.0587	.1084	.2184	.5803			
	3	$V(\mu^*)$				.0874	.0914	.0993	.1164	.1584	.3078				
		$V(\sigma^*)$				.1112	.1399	.1836	.2574	.4073	.8639				
		$Cov(\mu^*, \sigma^*)$				.0000	.0107	.0294	.0648	.1442	.4054				
	4	$V(\mu^*)$					.0939	.1002	.1167	.1780					
		$V(\sigma^*)$					.1843	.2591	.4108	.8714					
		$Cov(\mu^*, \sigma^*)$					.0000	.0216	.0716	.2397					
	5	$V(\mu^*)$						.1032	.1168						
		$V(\sigma^*)$						.4118	.8747						
		$Cov(\mu^*, \sigma^*)$						.0000	.0793						
14	0	$V(\mu^*)$	.0714	.0725	.0743	.0772	.0816	.0883	.0988	.1155	.1435	.1938	.2950	.5419	1.4654
		$V(\sigma^*)$	.0395	.0444	.0500	.0568	.0650	.0753	.0886	.1065	.1318	.1702	.2350	.3666	.7687
		$Cov(\mu^*, \sigma^*)$	.0000	.0023	.0055	.0099	.0159	.0242	.0360	.0533	.0799	.1239	.2048	.3851	.9945
	1	$V(\mu^*)$		.0733	.0749	.0775	.0817	.0883	.0992	.1175	.1505	.2164	.3736	.9337	
		$V(\sigma^*)$		.0505	.0576	.0664	.0773	.0915	.1106	.1376	.1789	.2488	.3914	.8284	
		$Cov(\mu^*, \sigma^*)$		.0000	.0034	.0081	.0149	.0246	.0390	.0612	.0981	.1660	.3158	.8105	
	2	$V(\mu^*)$			.0762	.0784	.0821	.0884	.0993	.1193	.1594	.2543	.5858		
		$V(\sigma^*)$			.0668	.0781	.0929	.1127	.1408	.1836	.2561	.4038	.8547		
		$Cov(\mu^*, \sigma^*)$			.0000	.0050	.0124	.0236	.0411	.0704	.1243	.2427	.6293		
	3	$V(\mu^*)$				.0800	.0831	.0887	.0993	.1212	.1737	.3567			
		$V(\sigma^*)$				.0933	.1137	.1425	.1863	.2604	.4108	.8689			
		$Cov(\mu^*, \sigma^*)$				.0000	.0079	.0206	.0422	.0825	.1714	.4609			
	4	$V(\mu^*)$					.0852	.0896	.0994	.1237	.2104				
		$V(\sigma^*)$					.1430	.1875	.2626	.4148	.8770				
		$Cov(\mu^*, \sigma^*)$					.0000	.0141	.0411	.1020	.3022				
	5	$V(\mu^*)$						.0922	.0997	.1286					
		$V(\sigma^*)$						.2633	.4166	.8812					
		$Cov(\mu^*, \sigma^*)$						.0000	.0338	.1496					
	6	$V(\mu^*)$							.1022						
		$V(\sigma^*)$							.8825						
		$Cov(\mu^*, \sigma^*)$							.0000						

Table II. (Continued)

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10	11	12	13
15	0	$V(\mu^*)$	.0667	.0676	.0691	.0714	.0748	.0799	.0875	.0992	.1176	.1480	.2022	.3101	.5713	1.5404
		$V(\sigma^*)$	.0366	.0408	.0456	.0512	.0580	.0662	.0765	.0899	.1078	.1332	.1718	.2368	.3689	.7723
		$Cov(\mu^*, \sigma^*)$	.0000	.0019	.0046	.0082	.0130	.0195	.0284	.0409	.0590	.0868	.1325	.2163	.4020	1.0273
	1	$V(\mu^*)$		.0683	.0696	.0717	.0749	.0799	.0877	.1001	.1207	.1574	.2296	.4004	1.0037	
		$V(\sigma^*)$		.0460	.0519	.0590	.0678	.0787	.0929	.1121	.1393	.1807	.2508	.3939	.8322	
		$Cov(\mu^*, \sigma^*)$		.0000	.0028	.0067	.0120	.0194	.0299	.0453	.0689	.1079	.1791	.3355	.8496	
	2	$V(\mu^*)$			.0706	.0724	.0753	.0800	.0877	.1007	.1239	.1697	.2769	.6471		
		$V(\sigma^*)$			.0594	.0684	.0798	.0946	.1145	.1427	.1856	.2584	.4065	.8586		
		$Cov(\mu^*, \sigma^*)$			.0000	.0040	.0098	.0181	.0304	.0496	.0812	.1389	.2649	.6740		
	3	$V(\mu^*)$				.0738	.0762	.0804	.0877	.1011	.1279	.1908	.4069			
		$V(\sigma^*)$				.0802	.0953	.1157	.1446	.1886	.2629	.4138	.8731			
		$Cov(\mu^*, \sigma^*)$				.0000	.0061	.0153	.0298	.0541	.0987	.1962	.5112			
	4	$V(\mu^*)$					.0780	.0813	.0879	.1014	.1340	.2472				
		$V(\sigma^*)$					.1161	.1455	.1901	.2655	.4182	.8817				
		$Cov(\mu^*, \sigma^*)$					.0000	.0099	.0270	.0590	.1295	.3585				
	5	$V(\mu^*)$						.0835	.0886	.1016	.1487					
		$V(\sigma^*)$						.1906	.2667	.4205	.8866					
		$Cov(\mu^*, \sigma^*)$						.0000	.0196	.0644	.2126					
	6	$V(\mu^*)$							.0911	.1017						
		$V(\sigma^*)$							.4212	.8888						
		$Cov(\mu^*, \sigma^*)$							.0000	.0704						

This table is a continuation of Table II in [2]. The entries in Table I, as well as in Table II of this present paper, have been rounded to four decimal places for convenience. Readers who desire more precision may obtain copies of the original tables containing eight decimal places from the authors. The results in the eight-decimal-place table are exact to seven places but rounding may cause some of the figures in the eighth place to be a few units in error.

If the coefficients of an estimate are sought for a value of r_1 not given in the table, the same procedure can be followed as that mentioned in Part I of this series.

The variances of the estimates and their covariances are given in Table II in terms of σ^2. This table is a continuation of Table III in [2] and the results are given to only four decimal places for convenience.

Table III shows the efficiency of the estimates for every case of censoring relative to the corresponding estimate obtained by complete samples.

3. Alternative estimate. The alternative estimate was proposed by Gupta [1] to replace the best linear estimate when sample sizes are greater than 10 and censoring was from one side only. This estimate was generalized to the case of double censoring in Part I of this present series. The variance of the alternative estimates and their efficiencies relative to the best linear estimate for samples of sizes 12 and 15 under every case of censoring are given in Table IV.

The authors know of no instance where the alternative estimates have been compared previously for sample sizes this large.

4. Comments. The conclusions mentioned in [2], Section 5, hold true here and, in fact, appear much stronger for increasing sample size. Several points are worth emphasis:

(1) In estimation of the mean, the relative efficiency holds up—about 65 per cent or better—as long as the median value remains known. (For an even n, it is about 70 per cent or better as long as the two middle values are uncensored.) This result was anticipated because the asymptotic efficiency of the median is $2/\pi = 63.7\%$.

Another way of presenting this same finding can be seen clearly from Fig. 1 which shows the relative efficiency of the best linear estimate of μ under all conditions of censoring a sample of size 15 from the normal distribution. Each one of the curves shows the efficiency of the estimate of the mean for a certain number of known elements $[k = n - (r_1 + r_2)]$ for all possible values of r_1 and r_2. The efficiency attains its maximum whenever the middle element is known.

(2) From Table III, one can see that, for fixed values of censoring from one side, the efficiency of the estimate of the standard deviation decreases approximately in equal amounts with each increment in the number of censored elements on the opposite side.

This is illustrated by Fig. 2 which shows the relative efficiency of the best linear estimate of σ under all conditions of censoring a sample of size 15 from the

Table III

Percentage efficiencies of the estimate of the mean (μ^*) and standard deviation (σ^*) for censored samples relative to uncensored samples in a normal population for $11 \leq n \leq 15$
(Continuation of Table IV in reference [2].)

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10
11	0	μ^*	100.0	97.89	94.13	88.20	79.55	67.74	52.92	36.31	20.23	7.46	
		σ^*	100.0	86.02	73.43	61.82	51.04	40.95	31.51	22.68	14.45	6.85	
	1	μ^*		96.36	93.29	87.98	79.54	67.06	50.37	31.04	12.82		
		σ^*		72.55	60.53	49.55	39.44	30.10	21.47	13.54	6.35		
	2	μ^*			91.20	87.04	79.46	66.64	47.21	22.75			
		σ^*			49.14	38.86	29.48	20.93	13.16	6.16			
	3	μ^*				84.56	78.84	66.39	41.29				
		σ^*				29.31	20.71	12.98	6.07				
	4	μ^*					76.32	66.28					
		σ^*					12.93	6.04					

12	0	μ^*	100.0	98.16	94.95	90.04	82.98	73.38	61.13	46.72	31.45	17.33	6.39
		σ^*	100.0	87.18	75.57	64.84	54.80	45.39	36.53	28.18	20.33	12.98	6.17
	1	μ^*		96.78	94.13	89.74	82.97	73.14	59.83	43.52	26.08	10.60	
		σ^*		74.78	63.63	53.40	43.92	35.09	26.87	19.22	12.16	5.72	
	2	μ^*			92.25	88.75	82.75	73.08	58.63	39.40	18.06		
		σ^*			53.01	43.33	34.44	26.25	18.71	11.81	5.55		
	3	μ^*				86.49	81.89	73.08	57.21	31.85			
		σ^*				34.25	25.99	18.47	11.63	5.46			
	4	μ^*					79.43	72.84	54.68				
		σ^*					18.39	11.55	5.42				
	5	μ^*						70.91					
		σ^*						5.41					

Table III (continued)

n	r_1	r_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13
13	0	μ^*	100.0	98.36	95.60	91.44	85.55	77.61	67.43	55.13	41.39	27.48	15.03	5.55		
		σ^*	100.0	88.16	77.39	67.39	58.02	49.19	40.84	32.94	25.47	18.41	11.78	5.61		
	1	μ^*		97.11	94.80	91.08	85.51	77.55	66.80	53.27	37.78	22.19	8.93			
		σ^*		76.66	66.28	56.70	47.77	39.42	31.59	24.25	17.40	11.03	5.20			
	2	μ^*			93.10	90.09	85.17	77.55	66.38	51.28	33.10	14.69				
		σ^*			56.32	47.20	38.76	30.93	23.65	16.91	10.70	5.04				
	3	μ^*				88.03	84.20	77.43	66.10	48.58	24.99					
		σ^*				38.56	30.65	23.36	16.66	10.53	4.96					
	4	μ^*					81.88	76.80	65.93	43.21						
		σ^*					23.27	16.55	10.44	4.92						
	5	μ^*						74.54	65.86							
		σ^*						10.41	4.90							

14	0	μ^*	100.0	98.53	96.11	92.53	87.54	80.87	72.31	61.85	49.78	36.85	24.21	13.18	4.87
		σ^*	100.0	89.00	78.95	69.59	60.79	52.48	44.59	37.10	29.98	23.22	16.82	10.78	5.14
	1	μ^*		97.39	95.34	92.14	87.45	80.86	72.03	60.79	47.45	33.01	19.12	7.65	
		σ^*		78.28	68.56	59.55	51.12	43.20	35.74	28.71	22.09	15.88	10.09	4.77	
	2	μ^*			93.79	91.16	87.03	80.82	71.92	59.86	44.82	28.08	12.19		
		σ^*			59.19	50.57	42.55	35.06	28.06	21.52	15.43	9.79	4.62		
	3	μ^*				89.29	86.00	80.55	71.90	58.92	40.12	20.03			
		σ^*				42.35	34.76	27.73	21.21	15.18	9.62	4.55			
	4	μ^*					83.86	79.70	71.88	57.75	33.95				
		σ^*					27.63	21.08	15.05	9.53	4.51				
	5	μ^*						77.44	71.64	55.56					
		σ^*						15.01	9.49	4.48					
	6	μ^*							69.91						
		σ^*							4.48						

Table III (continued)

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10	11	12	13
15	0	μ^*	100.0	98.68	96.52	93.40	89.11	83.42	76.15	67.21	56.70	45.04	32.97	21.50	11.67	4.33
		σ^*	100.0	89.72	80.30	71.51	63.21	55.35	47.88	40.76	33.97	27.49	21.33	15.47	9.93	4.74
	1	μ^*		97.62	95.79	92.99	88.97	83.42	76.04	66.62	55.23	42.36	29.03	16.65	6.64	
		σ^*		79.69	70.54	62.04	54.06	46.54	39.42	32.68	26.30	20.28	14.60	9.30	4.40	
	2	μ^*			94.37	92.04	88.50	83.31	76.02	66.22	53.79	39.28	24.07	10.30		
		σ^*			61.70	53.53	45.89	38.73	32.00	25.68	19.73	14.18	9.01	4.27		
	3	μ^*				90.33	87.46	82.93	76.01	65.93	52.12	34.94	16.38			
		σ^*				45.70	38.43	31.66	25.33	19.43	13.93	8.85	4.20			
	4	μ^*					85.49	81.95	75.86	65.73	49.75	26.97				
		σ^*					31.55	25.18	19.27	13.80	8.76	4.15				
	5	μ^*						79.80	75.25	65.61	44.82					
		σ^*						19.22	13.74	8.71	4.13					
	6	μ^*							73.19	65.56						
		σ^*							8.70	4.12						

normal distribution. In this figure, the graphs for $r_1 = 0, 1, \cdots, 12$ show a parallelism as r_1 changes. Thus, for any corresponding value of r_2 the efficiency decreases by about the same amount for each change in the value of r_1.

(3) Using Table III again and reading the entries for σ^* in diagonal fashion, one can see that, for a given n and fixed uncensored sample size ($r_1 + r_2 =$ constant), the efficiency of the best estimate of σ is remarkably constant independently of how r_1 and r_2 are chosen. In other words, there is practically no difference in efficiency irrespective of the proportion of the relative censoring from either side.

This can be observed very clearly in Fig. 2. The approximate horizontal lines show constancy of the relative efficiencies of σ^* for the known elements (k) of the sample whatever may be the individual values of r_1 and r_2.

(4) From Table III (and graphs similar to Fig. 2), one can construct the following table showing how the efficiency in estimating σ^* varies with the number of uncensored values for each sample size to serve as a rough guide in censoring.

Rough guide for assessing approximate efficiency (per cent) of estimate of σ*

Sample Size n	Number of uncensored observations in sample, or $k = n - (r_1 + r_2)$													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	6	13	21	30	39	50	61	73	86	100				
12	6	12	19	27	35	44	54	64	75	87	100			
13	5	11	17	24	31	39	48	57	67	77	88	100		
14	5	10	15	22	28	36	43	51	59	69	78	89	100	
15	4	9	14	20	26	32	39	46	54	62	71	80	90	100

* These values are rounded averages of different combinations of censoring and are within 2 or 3 per cent in almost all cases.

The information in this rough guide for censoring is also illustrated in Fig. 3. The efficiency in estimating σ^* for varying proportions of censoring is shown in the graph for samples of size 10, 15, 20 and large n. The latter value was obtained from Gupta [1] and represented single censoring. However, as stated previously, the efficiency in estimating σ^* depends primarily upon the proportion of uncensored elements irrespective of the side and can be used in this way.

(5) Figs. 4 and 5 show the efficiencies of the alternative estimate from a sample of size 15 relative to the correspondingly best linear estimate for the mean and the standard deviation respectively.

Judging from these figures, the worst efficiencies of the alternative estimate for estimating both the mean and standard deviations are attained for singly censored samples (i.e., only r_1 or $r_2 = 0$). Thus, the alternative estimate is relatively more precise when applied to doubly censored samples. The alternative estimate was proposed by judging the results of a comparison of efficiencies using a singly

Table IV

Variances and Relative Efficiencies of Alternative Estimates of the Mean ($\mu^{*\prime}$) and Standard Deviation ($\sigma^{*\prime}$) for Censored Samples of Size 12 and 15 from a Normal Population

n	r_1	r_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13
12	0	($\mu^{*\prime}$) V.	.08333333	.08532138	.08921286	.09569251	.10617803	.12326618	.15197055	.20294226	.30233969	.53167098	1.30441131			
		R.E.	100.00	99.51	98.38	96.72	94.59	92.13	89.71	87.90	87.64	90.44	100.00			
		($\sigma^{*\prime}$) V.	.04693011	.05541733	.06584693	.07855768	.09459627	.11566732	.14474943	.18762545	.25730631	.39089620	.76011855			
		R.E.	99.89	97.03	94.20	92.04	90.42	89.29	88.66	88.66	89.60	92.36	100.00			
	1	($\mu^{*\prime}$) V.		.08688517	.08998634	.09535109	.10440203	.11990788	.14777036	.20261677	.33120159	.78642201				
		R.E.		99.11	98.38	97.32	96.20	95.03	94.25	94.50	96.48	100.00				
		($\sigma^{*\prime}$) V.		.06478228	.07715733	.09281482	.11337921	.14183822	.18414527	.25419403	.39403316	.81947829				
		R.E.		96.77	95.48	94.58	94.15	94.19	94.75	95.93	97.81	100.00				
	2	($\mu^{*\prime}$) V.			.09205743	.09613108	.10369458	.11794105	.14676609	.21537209	.46137626					
		R.E.			98.13	97.67	97.11	96.68	96.85	98.21	100.00					
		($\sigma^{*\prime}$) V.			.09299998	.11391167	.14281363	.18574083	.25691562	.39999265	.84501770					
		R.E.			95.09	94.97	95.30	96.14	97.51	99.28	100.00					
	3	($\mu^{*\prime}$) V.				.09854217	.10405259	.11637233	.14734277	.26167710						
		R.E.				97.77	97.79	97.99	98.85	100.00						
		($\sigma^{*\prime}$) V.				.14326687	.18682141	.25894808	.40403858	.85814724						
		R.E.				95.53	96.54	98.04	99.80	100.00						
	4	($\mu^{*\prime}$) V.					.10664702	.11546756	.15239128							
		R.E.					98.38	99.08	100.00							
		($\sigma^{*\prime}$) V.					.25964368	.40599317	.86468798							
		R.E.					98.17	99.98	100.00							
	5	($\mu^{*\prime}$) V.						.11751619								
		R.E.						100.00								
		($\sigma^{*\prime}$) V.						.86669069								
		R.E.						100.00								

15	0	$(\mu=1)$ V.	.06676667	.06783620	.06996743	.07323737	.07806417	.08512564	.09551426	.11105404	.13495703	.17330497	.23882285	.36232523	.63848625	1.54037885
		R.E.	100.00	99.60	98.71	97.46	95.84	93.88	91.66	89.32	87.12	85.40	84.66	85.58	89.47	100.00
		$(\sigma=1)$ V.	.03667132	.04199425	.04825549	.05545771	.06393382	.07415317	.08679542	.10290237	.12417377	.15360043	.19699885	.26744195	.40210996	.77230522
		R.E.	99.90	97.23	94.53	92.38	90.64	89.25	88.15	87.35	86.85	86.75	87.19	88.54	91.74	100.00
	1	$(\mu=1)$ V.		.06883025	.07060707	.07337808	.07754855	.08378381	.09319554	.10773719	.13110231	.17103630	.24649116	.41734709	1.00373313	
		R.E.		99.22	98.57	97.70	96.63	95.38	94.08	92.88	92.06	92.01	93.17	95.95	100.00	
		$(\sigma=1)$ V.		.04757932	.05459531	.06286426	.07280110	.08506272	.10067027	.12130831	.14998813	.19270808	.26346042	.40451706	.83217195	
		R.E.		96.62	95.12	93.93	93.08	92.54	92.31	92.40	92.86	93.75	95.21	97.38	100.00	
	2	$(\mu=1)$ V.			.07192272	.07413167	.07764224	.08314965	.09187723	.10612952	.13071264	.17732697	.28290628	.64711601		
		R.E.			98.23	97.70	97.02	96.23	95.45	94.86	94.81	95.72	97.89	100.00		
		$(\sigma=1)$ V.			.06297264	.07309736	.08556985	.10141243	.12232669	.15136196	.19461599	.26642265	.41072321	.85860349		
		R.E.			94.28	93.63	93.28	93.26	93.57	94.26	95.38	96.99	98.98	100.00		
	3	$(\mu=1)$ V.				.07566586	.07837611	.08298324	.09085279	.10482738	.13179671	.19342211	.40692760			
		R.E.				97.54	97.25	96.88	96.54	96.46	97.05	98.65	100.00			
		$(\sigma=1)$ V.				.08580912	.10194081	.12319669	.15265533	.19648232	.26920271	.41556039	.87312546			
		R.E.				93.42	93.50	93.93	94.73	95.97	97.66	99.58	100.00			
	4	$(\mu=1)$ V.					.08009668	.08351673	.09014086	.10361683	.13537627	.24716376				
		R.E.					97.36	97.40	97.49	97.89	98.99	100.00				
		$(\sigma=1)$ V.					.12350319	.15333514	.19764253	.27106519	.41878810	.88168661				
		R.E.					94.02	94.90	96.20	97.94	99.85	100.00				
	5	$(\mu=1)$ V.						.08536772	.09008179	.10250554	.14874275					
		R.E.						97.86	98.35	99.13	100.00					
		$(\sigma=1)$ V.						.19802892	.27197904	.42061691	.88658942					
		R.E.						96.26	98.05	99.97	100.00					
	6	$(\mu=1)$ V.							.09184730	.10169465						
		R.E.							99.17	100.00						
		$(\sigma=1)$ V.							.42120833	.88884100						
		R.E.							100.00	100.00						

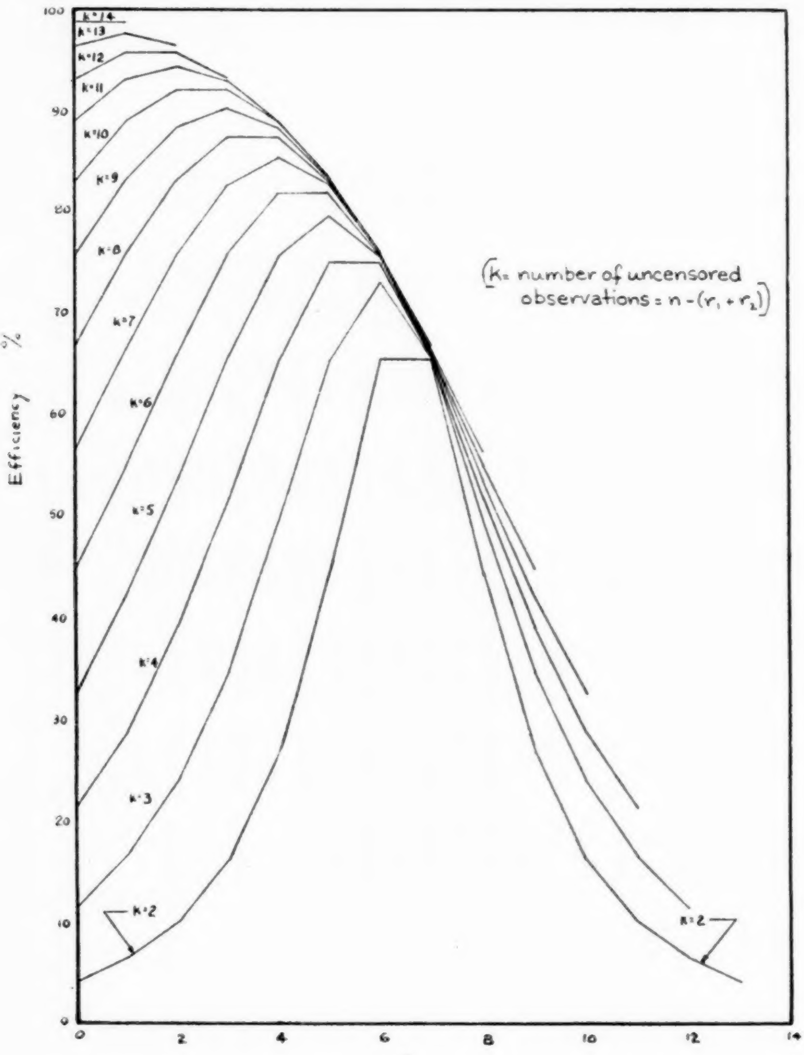


FIG. 1. Relative efficiency of the best linear estimate of μ under all conditions of censoring a sample of size 15 from the normal distribution.

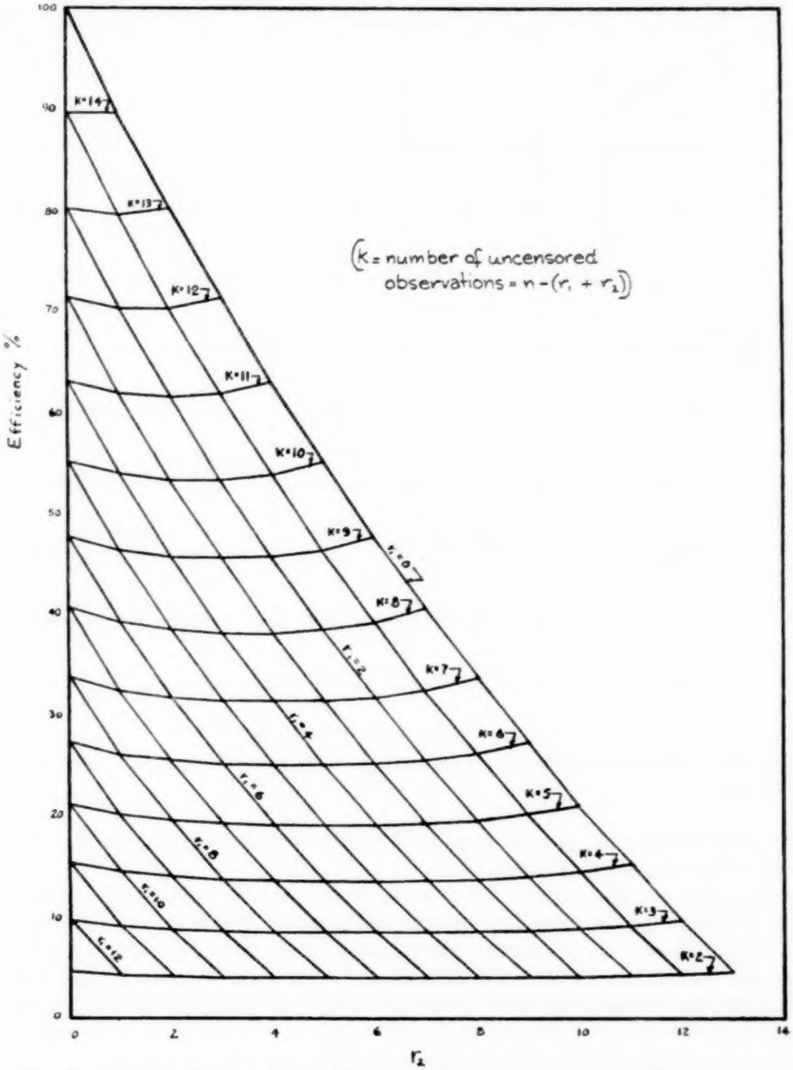


FIG. 2. Relative efficiency of the best linear estimate of σ under all conditions of censoring a sample of size 15 from the normal distribution.

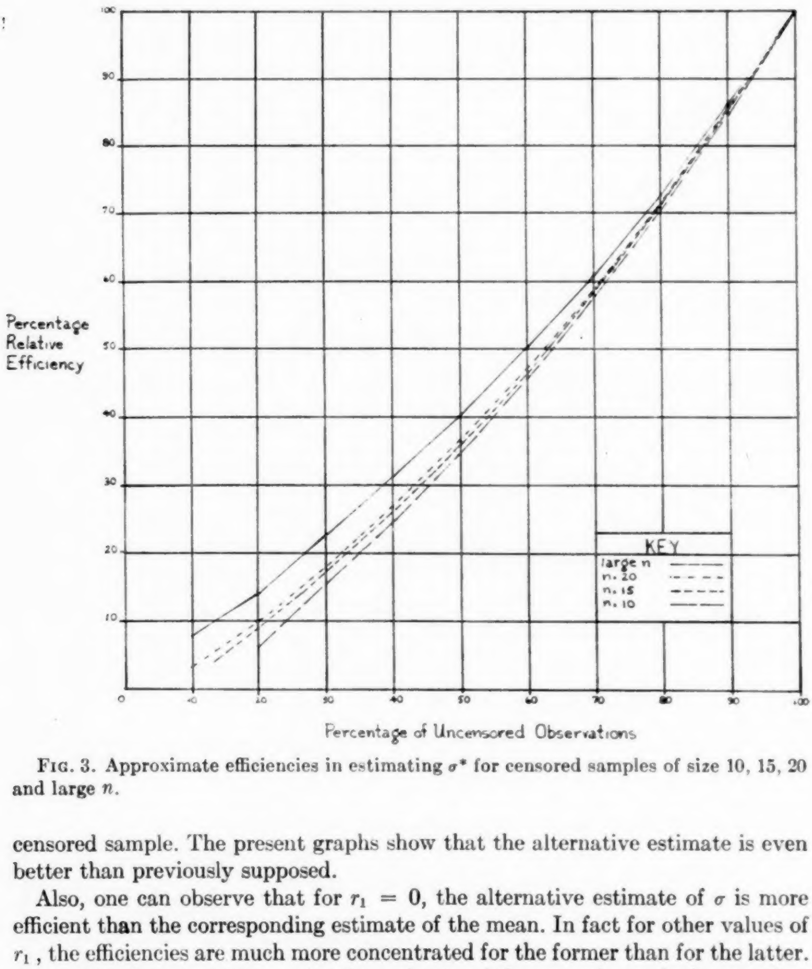


FIG. 3. Approximate efficiencies in estimating σ^* for censored samples of size 10, 15, 20 and large n.

censored sample. The present graphs show that the alternative estimate is even better than previously supposed.

Also, one can observe that for $r_1 = 0$, the alternative estimate of σ is more efficient than the corresponding estimate of the mean. In fact for other values of r_1, the efficiencies are much more concentrated for the former than for the latter. Again, the drop in efficiency for the estimate of the mean is much sharper than that for the standard deviation. This shows that the alternative estimate also appears better if one judges its value by considering its efficiency in estimation of the standard deviation rather than of the mean alone.

Addendum. The extension of Tables I, II, and III of this paper are now available for $16 \leqq n \leqq 20$ in eight decimal places. Copies may be obtained from the authors.

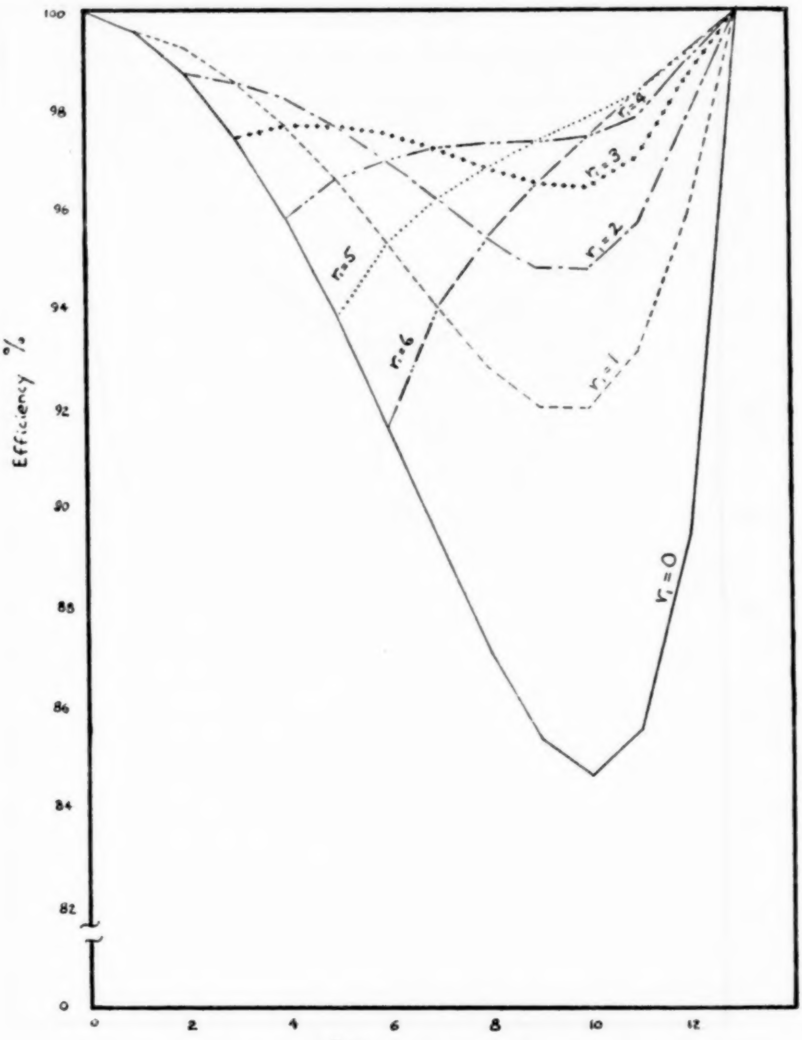


FIG. 4. Relative efficiency of the alternative estimate of μ under conditions of censoring a sample of size 15 from the normal population.

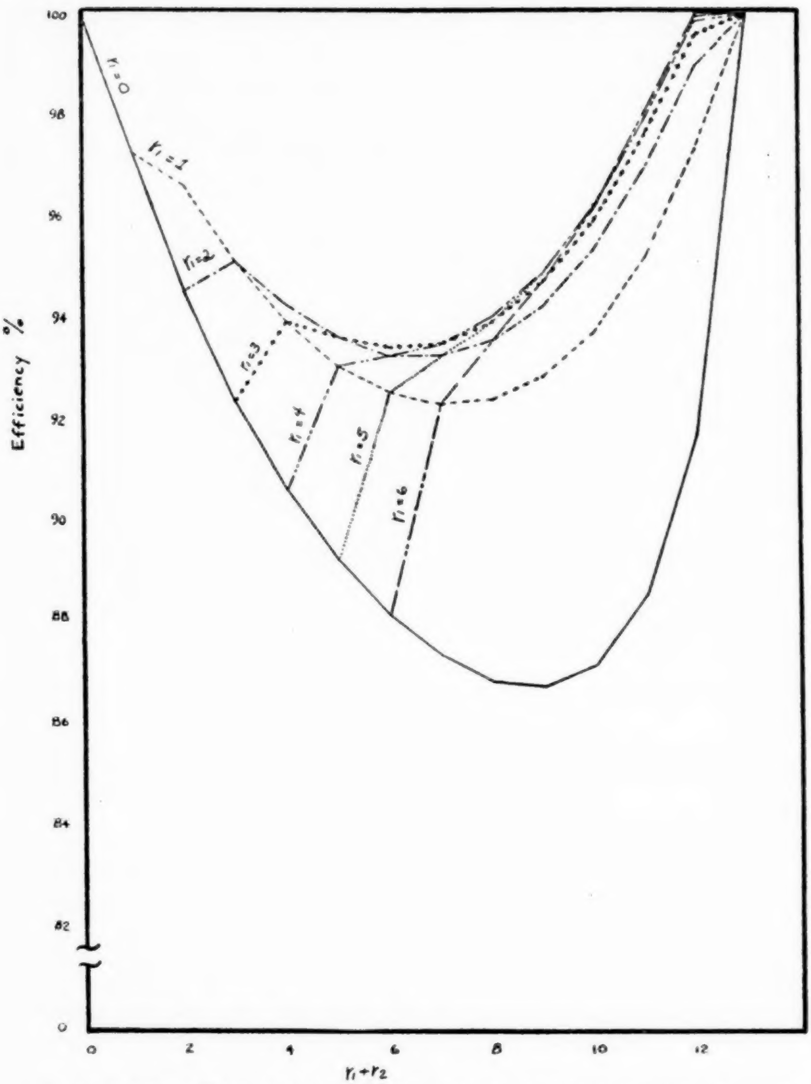


FIG. 5. Relative efficiency of the alternative estimate of σ under conditions of censoring a sample of size 15 from the normal population.

REFERENCES

[1] A. K. Gupta, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 88–95.

[2] A. E. Sarhan and B. G. Greenberg, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I. The normal distribution up to samples of size 10," *Ann. Math. Stat.*, Vol. 27, No. 2 (1956) pp. 427–451.

GENERALIZATIONS OF A GAUSSIAN THEOREM

BY PAUL S. DWYER

University of Michigan

1. Introduction and Summary. Plackett [1] has discussed the history and generalizations of the Gaussian theorem which states that least squares estimates are linear unbiased estimates with minimum variance. General forms of the theorem are due to Aitken [2], [3] and Rao [4], [5]. The essence of the proof for Aitken's general case consists in minimizing, simultaneously, certain quadratic forms involving linear combinations of the parameters. Plackett derived Aitken's result by using a matrix relation. The proof of the theorem follows quickly once the relation is established. A somewhat similar but simpler matrix relation is used by Rao ([4], page 10).

Aitken [2] and Rao [4], [5] obtain minimum variance with the use of Lagrange multipliers. Unless one has a method of working with matrices of derivatives it seems necessary to differentiate with respect to the many scalars constituting the matrices and to assemble the results in desired matrix form. Authors frequently give only the assembled results ([4], page 10, [5], page 17, [6], page 83).

The question arises as to whether it is possible to use the logically preferable matrix derivative methods of minimization. It is shown below that the use of matrices of partial derivatives [7] leads logically to the solution without the necessity of changing to and from scalar notation, or without the necessity of establishing some relation which implicitly contains the solution. Matrix derivative methods seem to be preferable methods for undertaking solutions of problems of simultaneous matrix minimization with side conditions for the same reason that derivative methods are preferable to the use of some (unknown) relation in solving problems of minimization involving scalars. They may also be used in establishing the relation which may then be verified without their use.

The paper includes generalizations of the results of Aitken [2], [3], Rao [4], [5], and David and Neyman [8]. It gives a general formula for simultaneous unbiased estimators of linear functions of parameters when the parameters are subject to linear restrictions and shows how the results are applicable to special cases. It provides formulas for the variance matrix of these estimators. It generalizes a matrix relation used by Plackett [1]. It uses the matrix square root transformation in establishing the general result for the variance of (weighted) residuals when there may be linear restrictions on the parameters. It provides a generalization of a formula of David and Neyman [8] in estimating the variance matrix of the unbiased linear estimators.

2. The least squares solution. The (inconsistent) observational equations are

$$A\theta = x \tag{2.1}$$

Received February 18, 1957; revised September 17, 1957.

and the true linear regression is given by

$$\mathcal{E}(x) = A\theta, \tag{2.2}$$

where the values of x, A, and θ are real. We set

$$A\theta - x = \epsilon \tag{2.3}$$

so that

$$\mathcal{E}(\epsilon) = 0. \tag{2.4}$$

In determining the least squares regression we have $\theta(s \times 1)$ as the vector of unknown parameters, $x(n \times 1)$ as the vector of measurements of the variable of regression, $\epsilon(n \times 1)$ as the vector of errors and $A(n \times s)$ as the matrix of measurements of the regressed variables. We take $s < n$ and A of rank s. Further, under the usual regression condition of fixed A,

$$\begin{aligned} V &= \mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T) = \mathcal{E}(\epsilon\epsilon^T) = V^T \\ &= \text{var}\,(x) = \text{var}\,\epsilon \end{aligned} \tag{2.5}$$

is the dispersion matrix of x and ϵ. We limit our discussion to the case where V is positive definite. A common dimensionless generalization of the least squares concept uses weights for the observations with $W = V^{-1}$ and leads to

$$Q = \epsilon^T V^{-1}\epsilon = (A\theta - x)^T V^{-1}(A\theta - x) \tag{2.6}$$

as the form to be minimized. The value of θ which minimizes (2.6) is known to be

$$\theta^* = (A^T V^{-1} A)^{-1} A^T V^{-1} x. \tag{2.7}$$

This result can be derived using symbolic matrix derivatives ([7], page 524). We have successively

$$Q = \theta^T A^T V^{-1} A\theta - \theta^T A^T V^{-1} x - x^T V^{-1} A\theta + x^T V^{-1} x, \tag{2.8}$$

$$\frac{\partial Q}{\partial\langle\theta\rangle} = J^T A^T V^{-1} A\theta + \theta^T A^T V^{-1} AJ - J^T A^T V^{-1} x - x^T V^{-1} AJ, \tag{2.9}$$

$$\frac{\partial\langle Q\rangle}{\partial\theta} = A^T V^{-1} A\theta K^T + A^T V^{-1} A\theta K - A^T V^{-1} xK^T - A^T V^{-1} xK, \tag{2.10}$$

and since Q is scalar, $K = K^T = 1$. Setting $\theta = \theta^*$ when $\dfrac{\partial\langle Q\rangle}{\partial\theta} = 0$, we get (2.7).

We note that θ^* is an unbiased estimate of θ because of (2.2) and (2.7).

3. Linear estimates with minimum variance. Now consider the k linear parametric functions

$$\phi = L\theta, \tag{3.1}$$

where $L = L(k \times s)$ is known. Then $\phi = \phi(k \times 1)$. We wish to find

$$\phi^* = L\theta^* \tag{3.2}$$

such that ϕ^* is an unbiased estimate of ϕ with minimum variance. This means that the diagonal terms of var ϕ (a matrix of order $k \times k$) should attain their minimum values simultaneously. Following Aitken we consider solutions of the form

$$\phi^* = Bx \tag{3.3}$$

and determine $B = B(k \times n)$. Rao [4] has shown that this homogeneous form is the general form. The relation

$$(BA - L)\theta = 0 \tag{3.4}$$

follows from (3.3), (2.2), and (3.1) in accordance with the requirement that ϕ^* be an unbiased estimate of ϕ.

Aitken [3] has shown using Lagrangian multipliers and Plackett [1] using a matrix relation that the value of θ^* in (3.2) which minimizes the diagonal term of var Q^* is identical with the θ^* resulting from least squares as given by (2.7). This Aitken theorem is a generalization of the Gaussian theorem that least squares linear estimators are unbiased with minimum variance.

Rao [5] further generalized the theorem with a consideration of linear restrictions on the parameters when $k = 1$. The argument is given below for the more general k. The preparation of the problem for minimization is similar to that of Rao in the special case with $k = 1$, though there are some modifications. The $u < s$ independent linear restrictions may be indicated by

$$g = R\theta \equiv 0, \tag{3.5}$$

where $R = R(u \times s)$ and $g = g(u \times 1)$, without loss of generality since any term not having some θ_i as a factor may be multiplied by $\theta_0 = 1$ and s replaced by $s' = s + 1$. We premultiply by the undetermined $D = D(k \times u)$ to get

$$DR\theta = Dg, \tag{3.6}$$

in which the matrix coefficient of θ has the same order as BA and L. Then the condition for unbiased estimation *and* the specific side conditions are incorporated in the matrix relation

$$(L - BA)\,\theta \equiv 0 \equiv DR\theta - Dg \tag{3.7}$$

so that the conditions for estimation can be written in the form

$$(L - BA - DR) = 0 \quad \text{and} \quad Dg = 0. \tag{3.8}$$

Specifically our purpose is the minimization of the diagonal terms of var ϕ^* subject to (3.8). Now

$$\text{var}\,\phi^* = \mathcal{E}(\phi^*\phi^{*T}) - \mathcal{E}(\phi^*)\mathcal{E}(\phi^{*T}) = B[\mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T)]B^T = BVB^T \tag{3.9}$$

We can then use

$$\psi = BVB^T + 2(L - BA - DR)\,\Lambda + 2Dg\mu^T, \tag{3.10}$$

where $\psi = \psi\ (k \times k)$, $\Lambda = \Lambda\ (s \times k)$, and $\mu = \mu\ (k \times 1)$ and differentiate with respect to B and D. We have

$$\frac{\partial\psi}{\partial\langle B\rangle} = JVB^T + BVJ^T - 2JA\Lambda,$$

$$\frac{\partial\langle\psi\rangle}{\partial B} = KBV + K^TBV - 2K\Lambda^TA^T,$$

so that the critical value, for each and every diagonal term, is given by

$$BV = \Lambda^TA^T. \tag{3.11}$$

Again

$$\frac{\partial\psi}{\partial\langle D\rangle} = -2JR\Lambda + 2Jg\mu^T,$$

$$\frac{\partial\langle\psi\rangle}{\partial D} = -2K\Lambda^TR^T + 2K\mu g^T,$$

$$\frac{\partial\psi_{ii}}{\partial D} = -2K_{ii}\Lambda^TR^T + 2K_{ii}\mu g^T,$$

so that, for each and every diagonal term

$$\Lambda^TR^T = \mu g^T,$$

so by (3.5),

$$\Lambda^TR^T = 0. \tag{3.12}$$

From (3.11) we get

$$B = \Lambda^TA^TV^{-1}. \tag{3.13}$$

Substituting in the first equation of (3.8), we arrive at

$$\Lambda^TA^TV^{-1}A + DR = L. \tag{3.14}$$

This equation and (3.12), for the special case with $k = 1$, were derived and emphasized by Rao [5], [17].

We next derive an estimate of ϕ in terms of Λ^T and θ^* for general k. We just multiply (3.14) by θ^* and use $R\theta^* = 0$ to get

$$\Lambda^TA^TV^{-1}A\theta^* = \phi^*. \tag{3.15}$$

The corresponding estimate in terms of Λ^T and x is

$$\phi^* = Bx = \Lambda^TA^TV^{-1}x. \tag{3.16}$$

It follows that θ^* satisfies

$$\Lambda^T A^T V^{-1} A\theta^* = \Lambda^T A^T V^{-1} x. \tag{3.17}$$

Equations (3.17) and (3.12) may be considered to be basic relations in θ^* and Λ^T.

4. The general Gaussian theorem. We next demonstrate the general Gaussian theorem that the value of θ^* obtained by least squares is consistent with that of (3.17) and (3.12). We note first that θ^* in the general solution is an $s \times 1$ vector and that the general solution is obtained by premultiplying θ^* by the fixed $k \times s$ matrix L. The general theorem is established by proving the typical case with $k = 1$ so that L, B, D, and Λ are vectors and (3.17) becomes

$$\lambda^T A^T V^{-1} A\theta^* = \lambda^T A^T V^{-1} x, \tag{4.1}$$

where λ^T is $\lambda^T(1 \times s)$. Also (3.12) becomes

$$\lambda^T R^T = 0. \tag{4.2}$$

Now we wish to minimize the scalar $Q = \epsilon^T V^{-1}\epsilon$, subject to the restriction conditions. Then

$$Q' = (A\theta - x)^T V^{-1}(A\theta - x) + 2(l - bA - dR)\lambda + 2\gamma R\theta. \tag{4.3}$$

Differentiation with respect to θ and d leads to the "normal" equations

$$A^T V^{-1} A\theta^* - A^T V^{-1} x + R^T \gamma^T = 0, \tag{4.4}$$

$$\lambda^T R^T = 0. \tag{4.5}$$

Premultiplying (4.4) by λ^T and substituting (4.5), we get

$$\lambda^T A^T V^{-1} A\theta^* = \lambda^T A^T V^{-1} x. \tag{4.6}$$

Since (4.6) and (4.5) are identical with (4.1) and (4.2), the λ's and θ's must be the same, so the general Gaussian theorem is true.

This solution, which is similar to that of Rao, is satisfactory in proving the generalized Gaussian theorem but it is not satisfactory in that it does not provide an explicit value of θ^* (only implicit relations involving the vector parameter λ) nor does it give an explicit expression for the unbiased linear estimator having minimum variance. These are provided in the sections following.

One further remark should be made before leaving these results on least squares The Eqs. (4.6) and (4.5) may be considered to be the normal equations of a general least squares problem expressed in terms of the vector parameter λ. Comparison of (4.6) with (2.7) shows that these normal equations can be obtained from the normal equations of the problem with no restrictions by premultiplication by λ^T where λ^T is subject to the conditions $\lambda^T R^T = 0$.

5. The explicit form of the estimator. It appears that no one has provided the explicit form for ϕ^* or for θ^*. Post multiplication of (3.14) by $(A^T V^{-1} A)^{-1} R^T$

followed by application of (3.12) eliminates Λ^T with the resulting

$$DR(A^TV^{-1}A)^{-1}R^T = L(A^TV^{-1}A)^{-1}R^T. \tag{5.1}$$

Now since $R(A^TV^{-1}A)^{-1}R^T$ is of order and rank u, we can write

$$D = L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}. \tag{5.2}$$

The value of Λ^T is then from (3.14)

$$\Lambda^T = L(A^TV^{-1}A)^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1} \tag{5.2}$$

and from (3.13),

$$\begin{aligned} B = {} & L(A^TV^{-1}A)^{-1}A^TV^{-1} \\ & - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1} \end{aligned} \tag{5.4}$$

so that

$$\begin{aligned} \phi^* = {} & L(A^TV^{-1}A)^{-1}A^TV^{-1}x \\ & - L(A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x \end{aligned} \tag{5.5}$$

is the linear unbiased estimator having minimum variance, and

$$\begin{aligned} \theta^* = {} & (A^TV^{-1}A)^{-1}A^TV^{-1}x \\ & - (A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x, \end{aligned} \tag{5.6}$$

and θ^* is the explicit solution of the normal equations. Rao did not give an explicit answer even for the case $k = 1$, since he did not derive an explicit formula for λ^T. The argument above covers the Rao case with L and λ vectors. Thus (5.5) and (5.6) hold with L a vector. As is pointed out above, the θ^* which results from least squares and from minimum variance is independent of L.

The results above are also general enough to include the Aitken results. These can be obtained formally from the above results by using the convention that $R^T[R(A^TV^{-1}A)R^T]^{-1}$ is 0 when $R = 0$, the formal equivalent of $u = 0$ side conditions. Thus the last terms drop from (5.5) and (5.6) for the Aitken problem.

The above results also generalize those of David and Neyman [8] who placed specifications on the dispersion matrix V. They defined V to be a diagonal matrix with

$$v_{ii} = \frac{\sigma^2}{P_{ii}}, \quad \text{where} \quad P_{ii} = \frac{\sigma^2}{\sigma_i^2}. \tag{5.7}$$

The formula for ϕ^* then becomes

$$\begin{aligned} \phi^* = {} & L(A^TPA)^{-1}A^TPx \\ & - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}A^TPx . \end{aligned} \tag{5.8}$$

Now B is ϕ^* with $x = I$, and θ^* is ϕ^* with $L = I$.

If $P_{ii} = \sigma P'_{ii}$ with $P'_{ii} = 1/\sigma_i^2$, we have

$$\phi^* = L(A^TP'A)^{-1}A^TP'x - L(A^TP'A)^{-1}R^T[R(A^TP'A)^{-1}R^T]^{-1}R(A^TP'A)^{-1}A^TP'x. \tag{5.9}$$

Then dropping the side conditions on the parameters we get

$$B = L(A^TPA)^{-1}A^TP = L(A^TP'A)^{-1}A^TP'. \tag{5.10}$$

When L is restricted to a vector, this is the David-Neyman result in matrix form.

When $V = I$, $L = I$ and $R = 0$ we have the common case of unweighted least squares regression

$$\phi^* = \theta^* = (A^TA)^{-1}A^Tx$$

and

$$B = (A^TA)^{-1}A^T. \tag{5.11}$$

The general results are immediately applicable to a variety of special cases involving specifications on V, specifications on L, and specifications on R. separately or in combinations.

6. The dispersion matrix of solutions. The dispersion matrix of solutions is var $\phi^* = BVB^T$. Using the value of B in (5.4), we get

$$\text{var}\ (\phi^*) = BVB^T = L(A^TV^{-1}A)^{-1}L^T - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}R(A^TV^{-1}A)^{-1}L^T. \tag{6.1}$$

When $k = 1$, this is an explicit result for the Rao problem. When there are no side conditions we have the Aitken result

$$\text{var}\ (\phi^*) = L(A^TV^{-1}A)^{-1}L^T. \tag{6.2}$$

When the values of x are uncorrelated with $v_{ii} = \sigma^2/P_{ii}$, (6.1) and (6.2) become

$$\text{var}\ (\phi^*) = L(A^TPA)^{-1}L^T\sigma^2 - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}L^T\sigma^2 \tag{6.3}$$

and

$$\text{var}\ (\phi^*) = L(A^TPA)^{-1}L^T\sigma^2. \tag{6.4}$$

When in addition the variables have a common variance σ^2, $\sigma_i^2 = \sigma^2$ and $P = I$. The Eqs. (6.3) and (6.4) appear with $(A^TA)^{-1}$ replacing $(A^TPA)^{-1}$.

If $\phi = \theta$, the above formulas appear with $L = I$. The simple case in which there are no side conditions, $\phi = \theta$, with variables uncorrelated but with equal variances gives

$$\text{var}\ (\theta^*) = (A^TA)^{-1}\sigma^2, \tag{6.5}$$

which is the formula for the dispersion matrix of regression coefficients in a common model.

7. Use of a matrix relation. The results (6.1) and (5.4) enable us to write a relation involving the value of B which gives the value of BVB^T having minimum diagnonal terms and the resulting matrix. In order to write this relation in compact form we use

$$C = (A^TV^{-1}A)^{-1} - (A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}R(A^TV^{-1}A)^{-1}, \tag{7.1}$$

which is Λ with $L = I$ to get

$$BVB^T = LCL^T + (B - LCA^TV^{-1})V(B - LCA^TV^{-1})^T. \tag{7.2}$$

The relation used by Plackett ([1], page 459) is a special case of this relation with the terms involving R deleted. Then $C = (A^TV^{-1}A)^{-1}$. Plackett's relation may be considered to be a generalization of the relation used by Gauss in establishing the theorem. Once the relation is established we see at once that the diagonal terms of BVB^T are minimized for general B when

$$B = LCA^TV^{-1} \tag{7.3}$$

as indicated in (5.4) and that the minimum values of the diagonal terms of the dispersion matrix are the diagonal terms of

$$BVB^T = LCL^T \tag{7.4}$$

as given in (6.1).

Once this general relation (7.2) is proposed, it may be verified by direct expansion. Then the whole solution of the problem of the minimization of the diagonal terms of the dispersion matrix of the estimators is immediately available as indicated by Plackett. If the relation is not known, and it has not been known previously for the general problem, it can be established with the use of matrix derivatives as shown above.

The various special cases of the general matrix relation result from the application of specified conditions to V, L, and R.

8. The variance of the residuals. Returning to the problem of least squares, we call $\mathcal{E}(\epsilon^TV^{-1}\epsilon)$ the variance of the (weighted) residuals. Then ϵ can be written

$$\epsilon = (ACA^TV^{-1} - I)x, \tag{8.1}$$

where C is given by (7.1), and ACA^TV^{-1}, and hence $ACA^TV^{-1} - I$, are idempotent. Hence

$$\begin{aligned} \epsilon^TV^{-1}\epsilon &= x^T(ACA^TV^{-1} - I)^TV^{-1}(ACA^TV^{-1} - I)x \\ &= x^TV^{-1}x - x^TV^{-1}ACA^TV^{-1}x. \end{aligned} \tag{8.2}$$

There is no loss in generality, for purpose of derivation, in assuming that x in (8.1) and (8.2) is a deviate with var $(x) = \mathcal{E}(xx^T) = V$.

For the Aitken problem, $C = (A^TV^{-1}A)^{-1}$ and we have

$$\epsilon^TV^{-1}\epsilon = x^TV^{-1}x - x^TV^{-1}A(A^TV^{-1}A)^{-1}A^TV^{-1}x. \tag{8.3}$$

To this we apply the triangular matrix square root transformation[1]

$$y = Wx \text{ with } W^TW = V^{-1}. \tag{8.4}$$

We then have

$$e^TV^{-1}e = y^Ty - y^TWA(A^TV^{-1}A)^{-1}A^TW^Ty \tag{8.5}$$

with

$$\text{var }(y) = E(Wxx^TW^T) = WVW^T = I_n \tag{8.6}$$

so that, using the trace

$$E(x^TV^{-1}x) = E(y^Ty) = n. \tag{8.7}$$

In order to find the expected value of the second term on the right in (8.5), we use the additional transformation

$$z = Sy \text{ with } S^TS = WA(A^TV^{-1}A)^{-1}A^TW^T, \tag{8.8}$$

where S is a triangular matrix. Since the rank of $WA(A^TV^{-1}A)^{-1}A^TW^T$ is s, S is of rank s, and there are $n - s$ rows identically zero. Then

$$e^TV^{-1}e = y^Ty - z^Tz, \tag{8.9}$$

and since

$$E(zz^T) = \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix}, \tag{8.10}$$

then

$$E(z^Tz) = s$$

and

$$E(e^TV^{-1}e) = E(y^Ty) - E(z^Tz) = n - s. \tag{8.11}$$

In the general problem with more complex C we have the additional quadratic form

$$x^TV^{-1}A(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^TR^T]^{-1}R(A^TV^{-1}A)^{-1}\cdot A^TV^{-1}x \tag{8.12}$$

whose matrix is of rank u. Application of (8.4) followed by application of

$$t = U_y, \text{ where } U^TU = WA(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} \cdot R(A^TV^{-1}A)^{-1}A^TW^T \tag{8.13}$$

[1] A triangular matrix square root, as applied to this problem, is a triangular matrix W defined by $W^TW = V^{-1}$. This should not be confused with the (non-triangular) algebraic matrix square root defined by $(V^{-\frac{1}{2}})^2 = V^{-1}$.

reduces this term to

$$t^T t \text{ with } E(t^T t) = u. \tag{8.14}$$

Then

$$E(\epsilon^T V^{-1} \epsilon) = n - s + u = n - (s - u). \tag{8.15}$$

This result is what one would expect. If the values of x were distributed normally, the positive definite quadratic form $\epsilon^T V^{-1} \epsilon$ would be distributed as χ^2 with $E(X^2) = n - (s - u)$ indicates the number of *independent* parameters.

This result is independent of k. In the Rao problem, $k = 1$, and the value of $E(\epsilon^T V^{-1} \epsilon)$ is $n - s + u$ as above. For the Aitken problem, $u = 0$, and the value is $n - s$. Where $V^{-1} = P/\sigma^2$ we have

$$E(\epsilon^T P \epsilon) = (n - s + u)\sigma^2 \tag{8.16}$$

and when $u = 0$, this is

$$E(\epsilon^T P \epsilon) = (n - s)\sigma^2 \tag{8.17}$$

as shown by David and Neyman for the case of uncorrelated variables ([8], pages 110–112). When $P = I$ this becomes

$$E(\epsilon^T \epsilon) = (n - s)\sigma^2 \tag{8.18}$$

as shown by Aitken using the properties of idempotent matrices ([3], page 139).

9. An estimator of the dispersion matrix of ϕ^*. David and Neyman [8] have provided an unbiased estimate of var ϕ^* for the case in which $V^{-1} = P/\sigma^2$, the x's are uncorrelated and L is a vector. A generalization related to the David-Neyman formula for the general problem is, for known V,

$$E^{-1} \operatorname{var} (\phi^*) = \frac{\epsilon^T V^{-1} \epsilon}{n - s + u} LCL^T, \tag{9.1}$$

since its expected value is the dispersion matrix of ϕ^*.

When V is known this formula is of little value since BVB^T can be computed and no estimation is necessary. However if V is not known, but P is, we have

$$\begin{aligned} E^{-1} \operatorname{var} (\phi^*) &= \frac{\epsilon^T P \epsilon}{n - s + u} \\ &\quad \cdot L\{(A^T PA)^{-1} - (A^T PA)^{-1} R^T [R(A^T PA)^{-1} R^T]^{-1} R(A^T PA)^{-1}\} L^T. \end{aligned} \tag{9.2}$$

When $P = I$, the case of equal variances, we have the important

$$\begin{aligned} E^{-1} \operatorname{var} (\phi^*) &= \frac{\epsilon^T \epsilon}{n - s + u} L\{(A^T A)^{-1} \\ &\quad - (A^T A)^{-1} R^T [R(A^T A)^{-1} R^T]^{-1} R(A^T A)^{-1}\} L^T. \end{aligned} \tag{9.3}$$

In the case of no side conditions we have

$$E^{-1}\text{ var }(\phi^*) = \frac{\epsilon^T P\epsilon}{n - s} L(A^T PA)^{-1}L^T. \tag{9.4}$$

Using the value of B in (5.10) we get

$$E^{-1}\text{ var }(\phi^*) = \frac{\epsilon^T P\epsilon}{n - s} BP^{-1}B^T. \tag{9.5}$$

If L is a vector, the estimate is a scalar. In the David-Neyman scalar notation, with the x's uncorrelated and B a row vector (λ) we have

$$\mu_1^2 = \frac{S_0}{n - s}\sum_{i=1}^{n}\frac{\lambda_i^2}{Pi}, \tag{9.6}$$

where $[\lambda_i] = \lambda = L(A^T PA)^{-1}A^T P = B$. Hence (9.2) gives the estimator matrix of var (ϕ^*) for a more general problem than does (9.6).

Appendix Showing Orders of Matrices and Conditions

Matrix	Order	Matrix	Order
X	$n \times 1$	ψ	$k \times k$
A	$n \times s$	Λ	$s \times k$
θ and θ^*	$s \times 1$	λ	$s \times 1$
ϵ	$n \times 1$	R	$u \times s$
V	$n \times n$	g	$u \times 1$
Q and Q'	1×1	D	$k \times u$
$A^T V^{-1}A$	$s \times s$	μ	$k \times 1$
L	$k \times s$	γ	$1 \times u$
ϕ and ϕ^*	$k \times 1$	$R(A^T V^{-1}A)^{-1}R^T$	$u \times u$
B	$k \times n$	C	$s \times s$
var ϕ^*	$k \times k$	P	$n \times n$
BVB^T	$k \times k$		

$u < s < n$, $u = 0$ gives Aitken problem, $k = 1$ gives Rao problem, $V^{-1} = P/\sigma^2$ gives David-Neyman condition.

REFERENCES

[1] R. L. PLACKETT, "A historical note on the method of least squares," *Biometrika*, Vol. 36 (1949), pp. 458–460.

[2] A. C. AITKEN, "On least squares and linear combinations of observations," *Proc. Roy. Soc. Edin.*, Vol. 55 (1934), pp. 42–48.

[3] A. C. AITKEN, "Studies in mathematics. IV. On linear approximations by least squares," *Proc. Roy. Soc. Edin.*, Vol. 62 (1945), pp. 138–146.

[4] C. R. RAO, "Generalization of Markoff's theorem and tests of linear hypotheses," *Sankhya*, Vol. 7 (1945), pp. 9–16.

[5] C. R. RAO, "Markoff's Theorem with linear restrictions on the parameters," *Sankhya*, Vol. 7 (1945), pp. 16–19.
[6] P. S. DWYER, "A matrix presentation of least squares and correlation theory," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 82–89.
[7] P. S. DWYER AND M. S. MACPHAIL, "Symbolic matrix derivatives," *Ann Math. Stat.*, Vol. 19 (1948), pp. 517–534.
[8] F. N. DAVID AND J. NEYMAN, "Extension of the Markoff Theorem on least squares," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 105–116.

A CENTRAL LIMIT THEOREM FOR SUMS OF INTERCHANGEABLE RANDOM VARIABLES[1]

By H. Chernoff and H. Teicher

Stanford University and Purdue University

1. Summary. A collection of random variables is defined to be interchangeable if every finite subcollection has a joint distribution which is a symmetric function of its arguments.

Double sequences of random variables X_{nk}, $k = 1, 2, \cdots, k_n$ $(\to \infty)$, $n = 1, 2, \cdots$, interchangeable (as opposed to independent) within rows, are considered. For each n, $X_{n1}, \cdots, X_{n,k_n}$ may (a) have a non-random sum, or (b) be embeddable in an infinite sequence of interchangeable random variables, or (c) neither. In case (a), a theorem is obtained providing conditions under which the partial sums have a limiting normal distribution. Applications to such well-known examples as ranks and percentiles are exhibited. Case (b) is treated elsewhere while case (c) remains open.

2. Terminology, notation and preliminaries. If X_n is a sequence of r.v.'s converging in probability (in measure) to a r.v. X, that is,

$$\lim P\{| X_n - X | > \epsilon\} = 0, \qquad \text{all } \epsilon > 0,$$

we abbreviate this by writing $X_n \xrightarrow{P} X$. This, in turn, implies $g(X_n) \xrightarrow{P} g(X)$ for any continuous function $g(x)$. If the corresponding c.d.f.'s $F_{X_n}(x) \to F_X(x)$ at all continuity points of the latter (in the sense of convergence of real numbers), we say X_n converges in law (or distribution) to X and write $X_n \xrightarrow{L} X$. We shall use frequently without ado the facts that if $X_n \xrightarrow[P]{L} X$ and c_n is a sequence of positive constants such that $c_n \to c$, then $c_nX_n \xrightarrow[P]{L} cX$[3].

The notation $P\{A \mid B\}$ will be used to designate the probability of an event A, given the occurrence of the event B, i.e., the conditional probability of A given B.

We shall be interested in and deal exclusively with r.v.'s whose joint c.d.f. is a symmetric function of its arguments. The same will then be true of the joint Fourier transform or characteristic function. This characteristic may also be expressed by stating that the joint distribution of $X_1, \cdots, X_k$ is invariant under permutations of the subscripts of the X's. Such random variables seem to have been introduced by de Finetti (cf. [4]). They have been termed "symmetrically dependent" by E. Sparre Anderson who also has studied some of their properties in a series of papers [1], [2]. By a quirk of terminology not in-

Received February 22, 1957; revised August 7, 1957.

[1] Research under contract with the Office of Naval Research.

frequent in mathematics, independent identically distributed r.v.'s are then subsumed under the category "symmetric dependence." On the grounds of brevity and connotation of the characteristic involved, we propose the sobriquet "interchangeable random variables" to denote any finite set of r.v.'s whose joint c.d.f. is symmetric.

In the case of an infinite sequence of r.v.'s, every finite subset of which has this property, Loève speaks of "exchangeable r.v.'s." However, the terminology of "interchangeability" of r.v.'s will be extended to include this case as well.

It is immediately evident that the r.v.'s, say $X_{i_1}, X_{i_2}, \cdots, X_{i_r}$, of any finite subcollection of a collection of interchangeable r.v.'s (i.r.v.'s) are themselves interchangeable and have a joint c.d.f. depending on r but not the permutation $(i_1, \cdots, i_r)$. In particular, the marginal c.d.f.'s $F_j(x) = F_{X_j}(x) = P\{X_j < x\}$ are identical for $j = 1, 2, \cdots, k$.

It is worth noting at the outset that it is, in general, not possible, to embed a given finite set of i.r.v.'s in an infinite set of i.r.v.'s (or even in a larger finite set). For example, if $P\{X_1 = 1, X_2 = 0\} = \frac{1}{2} = P\{X_1 = 0, X_2 = 1\}$ one cannot even adjoin a third r.v. so as to preserve interchangeability.

We commence with some elementary observations on the nature of i.r.v.'s. Two of these will be cast in the form of lemmas.

Suppose (as we shall throughout) that the i.r.v.'s under consideration have finite second order moments $EX_iX_j = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, dF_{X_i,X_j}(x, y), i, j = 1, 2, \cdots, k$. It is of course sufficient for this that when $m = 2$, $EX_1^m = \int_{-\infty}^{\infty} x^m\, dF_1(x) < \infty$. Take $\rho_{ii} = 1, i = 1, \cdots, k$ and define the (common) correlation coefficient between X_i and X_j by

$$\rho = \rho_{ij} = \frac{\operatorname{Cov}(X_i, X_j)}{\sqrt{\sigma_i^2\sigma_j^2}} = \frac{EX_iX_j - (EX_i)(EX_j)}{\sqrt{E(X_i - EX_i)^2E(X_j - EX_j)^2}}$$
$$= \frac{EX_1X_2 - (EX_1)^2}{E(X_1 - EX_1)^2}, \qquad i \neq j.$$

Then, the positive semi-definiteness of the correlation matrix

$$R = \{\rho_{ij}\} = \begin{pmatrix} 1\rho\cdots\rho \\ \rho 1\cdots\rho \\ \cdot \\ \quad\cdot \\ \qquad\cdot \\ \rho\rho\cdots I \end{pmatrix}$$

constrains ρ to be at least $-[1/(k-1)]$, where k is the number of i.r.v.'s. For if J is the $k \times k$ matrix consisting entirely of ones and I is the identity matrix of order k,

$$|R| = |\rho J + (1-\rho)I| = [k\rho + (1-\rho)][1-\rho]^{k-1} \geqq 0.$$

Thus, $\rho \geqq -[1/(k-1)]$. Consequently, the correlation between any pair of an infinite collection of interchangeable r.v.'s cannot be negative.

The following simple lemmas which we present without proof are useful. Let X and Y designate the vectors $(X_1, X_2, \cdots, X_k)$ and $(Y_1, Y_2, \cdots, Y_k)$.

LEMMA 1. *If $X_1, X_2, \cdots, X_k$ are interchangeable* r.v.'s *and $Y = \psi(X)$ is defined by $Y_j = \Phi[X_j, g(X)]$, $j = 1, 2, \cdots, k$, where Φ and g are Borel measurable functions, the latter being symmetric in its k arguments, then $Y_1, Y_2, \cdots, Y_k$ are interchangeable.*

LEMMA 2. *If $Y = (Y_1, Y_2, \cdots, Y_k)$ is a random permutation of the interchangeable* r.v.'s $X_1, X_2, \cdots, X_k$, *then Y has the same distribution as X.*

3. Background and Framework. The term "Central Limit Theorem" is a loose designation for one of an agglomeration of theorems dealing with limiting normality of distributions of sums of random variables—in the classical treatment—independent random variables.

The early results of De Moivre and Laplace have been succeeded by ever more powerful theorems set in an increasingly general framework. Recent works [5], [6] commence with a double sequence of rowwise *independent* r.v.'s (i.e., the r.v.'s within each row are independent)

$$\begin{array}{cccc} X_{11}, & X_{12}, & \cdots, & X_{1k_1} \\ X_{21}, & X_{22}, & \cdots, & X_{2k_2} \\ \vdots & \vdots & & \\ X_{n1}, & X_{n2}, & \cdots, & X_{nk_n} \\ \vdots & \vdots & \cdots & \vdots \end{array}$$

(where $k_n \to \infty$) and investigate the limiting distributions, i.e., c.d.f.'s of the row sums, say $S_n = \sum_{k=1}^{k_n} X_{nk}$. To render the problem more meaningful the r.v.'s are required to be "infinitesimal" (or asymptotically constant), i.e.,

$$\lim_{n\to\infty} \max_{1\leq i\leq k_n} P\{|X_{ni}| > \epsilon\} = 0, \qquad \text{all } \epsilon > 0.$$

A famous theorem of Khintchine asserts that the class of limiting distribution of such sums S_n coincides with the class of infinitely divisible laws [5]. A necessary and sufficient condition that the limiting distribution (assuming one exists) of sums of row-wise independent infinitesimal r.v.'s be normal is well known, namely, $\max_{1<i<k_n} |X_{ni}| \xrightarrow{P} 0$. (This actually implies infinitesimality here). For purposes of comparison with Theorem 1 of the next section we state the following result of Raikov (cf. [5]):

If Z_{nk}, $k = 1, \cdots, k_n$ are infinitesimal rowwise independent r.v.'s with zero means and finite variances σ_{nk}^2 with $\sum_{k=1}^{k_n} \sigma_{nk}^2 = 1$, a necessary and sufficient condition that the c.d.f. of $\sum_{k=1}^{k_n} Z_{nk}$ converges to the normal c.d.f. with mean 0 and variance 1 is that $\sum_{k=1}^{k_n} Z_{nk}^2 \xrightarrow{P} 1$.

Attempts have been made to relax the requirement of independence with varying degrees of success. Perhaps a natural and useful generalization is to double sequences of interchangeable random variables.

In this direction, let X_{ni}, $i = 1, \cdots, k_n$ comprise a (finite) set of i.r.v.'s for every $n = 1, 2, \cdots,$.

If we stipulate that $\lim_{n\to\infty} P\{|X_{n1}| > \epsilon\} = 0$, all $\epsilon > 0$, the question of the nature of the class C^* of all limiting distributions of row sums may again be posed. Clearly, C^* includes all stable distributions but contains others as well. This follows from a result of von Mises [7] who showed that the distribution of the number S_{n,r_n} of unoccupied cells in a random casting of r_n objects into n cells approaches that of the Poisson when $n, r_n \to \infty$ in a manner such that the expected number of vacancies is constant. If the expected proportion of vacancies converges to a constant, then Irving Weiss [9] has shown that the limiting distribution, suitably normalized, is normal. But $S_{n,r_n} = \sum_{i=1}^{n} X_{ni}$ where the X_{ni} are i.r.v.'s assuming the values one or zero (according as the ith cell is empty or not). Therefore, the Poisson distribution and in fact all infinitely divisible distributions belong to C^*.

In this paper, we consider only the case of limiting normal distributions and treat the first of the following two situations:

(a) For each $n = 1, 2, \cdots$, the i.r.v.'s X_{ni}, $i = 1, 2, \cdots, k_n$ have a non-random sum.

(b) For each $n = 1, 2, \cdots$, the i.r.v.'s X_{ni}, $i = 1, \cdots, k_n$ are embeddable in an infinite sequence of i.r.v.'s.[2]

These cases are mutually exclusive since if $\sum_{i=1}^{k_n} X_{ni} = C_n$, the covariance of any pair of i.r.v.'s equals $-[1/(k_n - 1)]$ multiplied by the common variance. But then their correlation is negative, which is precluded (under case b) by a prior remark.

4. I.R.V.'s whose sum is non-random. For each $n = 1, 2, \cdots$, let X'_{nk}, $k = 1, 2, \cdots, k_n(\to \infty)$ be i.r.v.'s with finite variance $\sigma'^2_{nk} = \sigma'^2_{n1} = E(X'_{n1} - EX'_{n1})^2$ and satisfying the linear constraint

$$\text{(i}'\text{)} \qquad \sum_{i=1}^{k_n} X'_{ni} = C_n .$$

Naturally, under such a proviso we must investigate partial rather than complete row sums.

If we define

$$X_{ni} = \frac{1}{\sigma'_{ni}}\left(X'_{ni} - \frac{C_n}{k_n}\right),$$

the X_{ni} are, by Lemma 1, i.r.v.'s satisfying the relationships

$$\text{(i)} \qquad \sum_{i=1}^{k_n} X_{ni} = 0, \qquad n = 1, 2, \cdots,$$

and

(ii) $EX^2_{ni} = \sigma^2_{X_{ni}} = 1$, $i = 1, 2, \cdots, k_n$ and all $n = 1, 2, \cdots$,

[2] The theorems obtained for the case of infinite sequences of i.r.v.'s overlap results of Professors Blum and Rosenblatt of Indiana University and will appear in a joint publication.

We suppose, therefore, without loss of generality, that for each

$$n = 1, 2, \cdots, \{X_{nk}\}, k = 1, \cdots, k_n(\to \infty)$$

are rowwise i.r.v.'s satisfying (i) and (ii) and possessing the joint c.d.f. $F_n(x_1, x_2, \cdots, x_{k_n})$. We have then

THEOREM 1. *For each* $n = 1, 2, \cdots$, *let* $\{X_{ni}\}$ *be interchangeable random variables satisfying* (i) *and* (ii). *If*

$$\max_{1 \leq k \leq k_n} \frac{|X_{nk}|}{\sqrt{k_n}} \xrightarrow{P} 0, \tag{1}$$

$$\frac{1}{k_n} \sum_{k=1}^{k_n} X_{nk}^2 \xrightarrow{P} 1, \tag{2}$$

and $m_n < k_n$ *is a sequence of positive integers such that* $\lim_{n\to\infty} m_n/k_n = \alpha$, $0 < \alpha < 1$, *then*

$$\lim_{n\to\infty} P\left\{\frac{1}{\sqrt{m_n}} \sum_{k=1}^{m_n} X_{nk} < x\right\} = \frac{1}{\sqrt{2\pi(1-\alpha)}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2(1-\alpha)}\right) dy.$$

PROOF. For any set of real numbers x_{ni}, if $\max_{1 \leq i \leq k_n} |x_{ni}|/\sqrt{k_n} = o(1)$ and $\lim_{n\to\infty} (1/k_n) \sum_1^{k_n} x_{ni}^2 = 1$, then

$$\max_{1 \leq i \leq k_n} \frac{|x_{ni}|}{\sqrt{\sum_1^{k_n} x_{ni}^2}} = o(1).$$

It follows directly that if the x_{ni} are r.v.'s and the analogous conditions are true in "probability" the conclusion holds "in probability." That is, (1) and (2) imply

$$\max_{1 \leq i \leq k_n} \frac{|X_{ni}|}{\sqrt{\sum_{i=1}^{k_n} X_{ni}^2}} \xrightarrow{P} 0. \tag{5}$$

Next, let $Y_{n1}, \cdots, Y_{nk_n}$ be a randomly selected permutation of $X_{n1}, \cdots, X_{nk_n}$. Then even when it is stipulated that X_{ni} = fixed real number x_{ni}, $i = 1, 2, \cdots, k_n$, the quantity

$$U_n = \left(\sum_{i=1}^{k_n} X_{ni}^2\right)^{-1/2} \sum_{i=1}^{m_n} Y_{ni}$$

is a random variable.

Suppose that for some c.d.f. $G(u)$ and arbitrary $\epsilon > 0$, there exists $\delta_\epsilon > 0$ and integral $N_1(\epsilon)$ (all independent of $x_{n1}, \cdots, x_{nk_n}$) such that

$$\max_{1 \leq i \leq k_n} \frac{|x_{ni}|}{\sqrt{\sum_1^{k_n} x_{ni}^2}} < \delta_\epsilon$$

implies

$$|P\{U_n < u | X_{ni} = x_{ni}, i = 1, \cdots, k_n\} - G(u)| < \epsilon \tag{6}$$

for all $n > N_1(\epsilon)$ and continuity points u of $G(u)$. By (5), there exists $N_2(\epsilon)$ such that for all $n > N_2(\epsilon)$, say

$$\epsilon > P\left\{\max_{1\leq i\leq k_n} \frac{|X_{ni}|}{\sqrt{\sum_1^{k_n} X_{ni}^2}} > \delta_\epsilon\right\} = P\{\bar{A}_n\}. \tag{7}$$

Then, from (6) and (7) for arbitrary $\epsilon > 0$ and $n > \max[N_1(\epsilon), N_2(\epsilon)]$ and continuity points u of $G(u)$,

$$\begin{aligned} &| P\{U_n < u\} - G(u) | \\ &= \left| \int_{R^{k_n}} [P\{U_n < u \mid X_{ni} = x_{ni}, i = 1, \cdots, k_n\} - G(u)]\, dF_n(x_1, \cdots, x_{k_n}) \right| \\ &\leqq \int_{A_n} | P\{U_n < u \mid X_{ni}, i = 1, \cdots, k_n\} - G(u) |\, dF_n + \int_{\bar{A}_n} dF_n \leqq 2\epsilon. \end{aligned} \tag{8}$$

For simplicity in writing, let Q be an r.v. with c.d.f. $G(u)$; then for $\lambda > 0$, $Q_\lambda = (1/\lambda)\, Q$ is an r.v. with distribution $G(\lambda u)$. Under the proviso (6), (8) shows that

$$U_n = \frac{\sum_{i=1}^{m_n} Y_{ni}}{\sqrt{\sum_1^{k_n} X_{ni}^2}} \xrightarrow{L} Q.$$

On the other hand, according to (2),

$$\sqrt{\frac{1}{k_n} \sum_{i=1}^{k_n} X_{ni}^2} \xrightarrow{P} 1.$$

Consequently, (see, e.g., [3]),

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} Y_{ni} = \sqrt{\frac{k_n}{m_n}}\, U_n \sqrt{\frac{1}{k_n} \sum_{i=1}^{k_n} X_{ni}^2} \xrightarrow{L} \frac{1}{\sqrt{\alpha}}\, Q = Q_{\sqrt{\alpha}}.$$

But by Lemma 2, $\sum_{i=1}^{m_n} Y_{ni}$ and $\sum_{i=1}^{m_n} X_{ni}$ have the same distribution, and thus, under the proviso (6),

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} X_{ni} \xrightarrow{L} Q_{\sqrt{\alpha}}\, . \tag{9}$$

It remains to verify (6) for $G(u)$ the c.d.f. of $Q = N_{o,\alpha(1-\alpha)}$, where N_{μ,σ^2} represents a normal random variable with mean μ and variance σ^2. To do so, it suffices to prove that

$$U_n = \left(\sum_{i=1}^{k_n} x_{ni}^2\right)^{-1/2} \sum_{i=1}^{m_n} Y_{ni} \xrightarrow{L} Q = N_{o,\alpha(1-\alpha)}\, , \tag{10}$$

providing $Y_{n1}, Y_{n2}, \cdots, Y_{nk_n}$ is a random permutation of the fixed real numbers

$x_{n1}, x_{n2}, \cdots, x_{nk_n}$, where

$$\max_{1 \leq k \leq k_n} \left(\sum_{i=1}^{k_n} x_{ni}^2 \right)^{-1/2} |x_{nk}| = o(1) \tag{11}$$

and $\sum_{i=1}^{k_n} x_{ni} = 0$. A theorem of Noether [8] states that the distribution of

$$L_n = \sum_{i=1}^{k_n} d_{ni} Y_{ni}$$

converges to the normal distribution (when normalized by its mean and standard deviation) if the d_{ni} are fixed real numbers such that

$$D_{n,s} = \frac{\frac{1}{k_n} \sum_{i=1}^{k_n} (d_{ni} - \bar{d}_n)^s}{\left[\frac{1}{k_n} \sum_{i=1}^{k_n} (d_{ni} - \bar{d}_n)^2 \right]^{s/2}} = O(1) \qquad \text{for } s = 3, 4, \cdots$$

and

$$A_{ns} = \frac{\sum_{i=1}^{k_n} (x_{ni} - \bar{x}_n)^s}{\left[\sum_{i=1}^{k_n} (x_{ni} - \bar{x}_{n.})^2 \right]^{s/2}} = o(1), \qquad \text{for } s = 3, 4, \cdots$$

with $\bar{d}_n = 1/k_n \sum_{i=1}^{k_n} d_{ni}$ and $\bar{x}_r = 1/k_n \sum_{i=1}^{k_n} x_{ni} = 0$. Let $d_{ni} = 1$ for $1 \leq i \leq m_n$ and 0 for $m_n + 1 \leq i \leq k_n$. Then $D_{n,s} = 0(1)$ for $s = 3, 4, \cdots$. Furthermore, from (11),

$$A_{ns} = \frac{\sum_{i=1}^{k_n} x_{ni}^s}{\left(\sum_{i=1}^{k_n} x_{ni}^2 \right)^{s/2}} \leq \frac{\max_{1 \leq i \leq k_n} |x_{ni}|^{s-2}}{\left(\sum_{i=1}^{k_n} x_{ni}^2 \right)^{s-2/2}} = o(1) \quad \text{for } s = 3, 4, \cdots.$$

Thus Noether's theorem applies to $L_n = \sum_{i=1}^{m_n} Y_{ni}$ whose mean and variance we shall show to be given by $\mu_n = 0$ and

$$\sigma_n^2 = \left(\sum_{i=1}^{k_n} x_{ni}^2 \right) \frac{m_n(k_n - m_n)}{k_n(k_n - 1)}. \tag{12}$$

Then we shall have $L_n/\sigma_n \xrightarrow{L} N_{0,1}$ and the desired result

$$U_n = \frac{L_n}{\sigma_n} \sqrt{\frac{m_n(k_n - m_n)}{k_n(k_n - 1)}} \xrightarrow{L} N_{0,\alpha(1-\alpha)}.$$

We now conclude by evaluating μ_n and σ_n.

$$E(Y_{ni}) = \sum_{a=1}^{k_n} x_{na}/k_n = 0, \qquad E(Y_{ni}^2) = \sum_{a=1}^{n} x_{na}^2/k_n,$$

$$E(Y_{ni} Y_{nj}) = \left(\sum_{a \neq b} x_{na} x_{nb}\right)/k_n(k_n - 1) = -\sum_{a=1}^{n} x_{na}^2/k_n(k_n - 1), \qquad i \neq j.$$

Hence

$$\mu_n = E\left(\sum_{i=1}^{m_n} Y_{ni}\right) = 0$$

and

$$\sigma_n^2 = E\left(\sum_{i=1}^{m_n} Y_{ni}\right)^2 = \left(\sum_{i=1}^{k_n} x_{ni}^2\right)\left(\frac{m_n}{k_n} - \frac{m_n(m_n - 1)}{k_n(k_n - 1)}\right),$$

which matches (12), concluding the proof.

Corollary 1. *For each positive integer $n = 1, 2, \cdots$, let $\{X_{ni}\}$, $i = 1, 2, \cdots, k_n(\rightarrow \infty)$ be i.r.v.'s satisfying* (i) *and* (ii). *If $m_n < k_n$ is a sequence of positive integers with $\lim_{n\to\infty} m_n/k_n = \alpha$, $0 < \alpha < 1$ and*

$$(3) \qquad E[X_{n1}^4] = o(k_n), \qquad \mathrm{Cov}(X_{n1}^2, X_{n2}^2) = o(1),$$

then the conclusion of the theorem holds.

Proof. For any $\eta > 0$,

$$P\left\{\max_{1\leq i\leq k_n} \frac{|X_{ni}|}{\sqrt{k_n}} > \eta\right\} = P\left\{\bigcup_{1}^{k_n}\left[\frac{|X_{ni}|}{\sqrt{k_n}} > \eta\right]\right\}$$

$$\leqq k_n P\left\{\frac{|X_{n1}|}{\sqrt{k_n}} > \eta\right\} \leqq \frac{k_n E|X_{n1}|^4}{k_n^2\eta^2} = o(1)$$

and

$$P\left\{\left|\frac{1}{k_n}\sum_{1}^{k_n}(X_{ni}^2 - 1)\right| > \eta\right\} \leqq \frac{E\left[\sum_{1}^{k_n}(X_{ni}^2 - 1)\right]^2}{k_n^2\eta^2}$$

$$= \frac{k_n E(X_{n1}^2 - 1)^2 + k_n(k_n - 1)\,\mathrm{Cov}\,(X_{n1}^2, X_{n2}^2)}{k_n^2\eta^2} = o(1).$$

Corollary 2. *For each $n = 1, 2, \cdots$, let $\{X'_{ni}\}$, $i = 1, 2, \cdots, k_n(\rightarrow \infty)$ be i.r.v.'s with $\sum_{i=1}^{k_n} X'_{ni} = C_n$ and $\sum_{i=1}^{k_n}(X'_{ni})^2 = D_n^2 > 0$. If*

$$\max_{1\leq i\leq k_n} \frac{|X'_{ni} - C_n/k_n|}{\sqrt{1/k_n(D_n^2 - C_n^2/k_n)}} \xrightarrow{P} 0,$$

then the conclusion of the theorem holds for $(1/\sqrt{m_n})\sum_{1}^{m_n} X_{ni}$, *where*

$$X_{ni} = \frac{X'_{ni} - C_n/k_n}{[(1/k_n)(D_n^2 - C_n^2/k_n)]^{\frac{1}{2}}}.$$

Proof. Condition (2) is certainly satisfied since $1/k_n\sum_{1}^{k_n} X_{ni}^2 = 1$.

Corollary 3. *For each $n = 1, 2, \cdots$, let $\{X_{ni}\}$, $i = 1, \cdots, k_n(\rightarrow \infty)$ be i.r.v.'s with $EX_{n1} = 0$, $EX_{n1}^2 = 1$ and $\bar{X}_n = 1/k_n\sum_{1}^{k_n} X_{nk}$. If the $\{X_{ni}\}$ satisfy* (1), (2), *and*

$$(4) \qquad E(X_{n1}X_{n2}) = o(1),$$

then

$$\lim_{n\to\infty} P\left\{\frac{1}{\sqrt{m_n}} \sum_1^{m_n} (X_{ni} - \bar{X}_n) < x\right\} = \frac{1}{\sqrt{2\pi(1-\alpha)}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2(1-\alpha)}\right) dy.$$

PROOF. Let

$$Y_{ni} = \sqrt{\frac{k_n}{k_n - 1}} \frac{X_{ni} - \bar{X}_n}{\sqrt{1 - EX_{n1}X_{n2}}} = a_n(X_{ni} - \bar{X}_n).$$

Then, applying Lemma 1 with $g(X) = \bar{X}$, it follows that the $\{Y_{ni}\}$ are i.r.v.'s. Further, $\sum_{i=1}^{k_n} Y_{ni} = 0$ and $EY_{ni}^2 = 1$. Since

$$0 \leqq \max_{1\leqq i\leqq k_n} \frac{|Y_{ni}|}{\sqrt{k_n}} \leqq \frac{2a_n}{\sqrt{k_n}} \max_{1\leqq i\leqq k_n} |X_{ni}|,$$

(1) and (4) imply

$$\max_{1\leqq i\leqq k_n} \frac{|Y_{ni}|}{\sqrt{k_n}} \xrightarrow{P} 0.$$

Next, for every $\epsilon > 0$,

$$P\left\{\frac{1}{k_n}\left|\sum_{i=1}^{k_n} X_{ni}\right| > \epsilon\right\} \leqq \frac{E\left(\sum_{i=1}^{k_n} X_{ni}\right)^2}{k_n^2 \epsilon^2} = (k_n\epsilon^2)^{-1} + \frac{(k_n - 1)}{k_n} \frac{EX_{n1}X_{n2}}{\epsilon^2} = o(1).$$

That is, $\bar{X}_n \xrightarrow{P} 0$. Thus,

$$\frac{1}{k_n}\sum_{i=1}^{k_n} Y_{ni}^2 = \frac{(k_n - 1)^{-1}}{(1 - EX_{n1}X_{n2})}\left[\sum_{j=1}^{k_n} X_{nj}^2 - k_n\bar{X}_n^2\right] \xrightarrow{P} 1.$$

A direct application of the theorem to the $\{Y_{ni}\}$ shows that

$$\sqrt{\frac{k_n}{(k_n - 1)} \frac{1}{(1 - EX_{n1}X_{n2})}} \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} (X_{ni} - \bar{X}_n) \xrightarrow{L} N_{0,1-\alpha},$$

which, in view of (4), implies that

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} (X_{ni} - \bar{X}_n) \xrightarrow{L} N_{0,1-\alpha}.$$

COROLLARY 4. *For each* $n = 1, 2, \cdots$, *let* $\{X_{ni}\}$, $i = 1, \cdots, k_n$, *be i.r.v.'s with* $EX_{n1} = 0$, $EX_{n1}^2 = 1$. *If* m_n *is a sequence of positive integers such that* $\lim_n m_n/k_n = \alpha$, $0 < \alpha < 1$, *and the* $\{X_{ni}\}$ *satisfy*

$$\text{(4')} \qquad \operatorname{Cov}(X_{n1}, X_{n2}) = \frac{-1}{k_n - 1} + o(k_n^{-1})$$

and either (3) *or* (1) *and* (2), *then*

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} X_{ni} \xrightarrow{L} N_{0,1-\alpha}.$$

PROOF. Since, as shown in the proof of Corollary 1, (3) implies (1) and (2), it suffices to suppose that the latter obtain. But (4′) clearly implies (4) whence, according to Corollary 3,

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} X_{ni} - \sqrt{m_n} \bar{X}_n \xrightarrow{L} N_{0,1-\alpha}.$$

However, for positive arbitrary positive,

$$\begin{aligned} P\{|\sqrt{m_n}\bar{X}_n| > \epsilon\} &\leqq \frac{m_n}{\epsilon^2 k_n^2} E\left(\sum_{i=1}^{k_n} X_{ni}\right)^2 \\ &= \frac{m_n}{\epsilon^2 k_n}\left[1 + (k_n - 1)\left\{\frac{-1}{k_n - 1} + o(k_n^{-1})\right\}\right] \\ &= \frac{m_n}{\epsilon^2 k_n} o(1) \end{aligned}$$

employing (4′). Thus, $\sqrt{m_n}\, \bar{X}_n \xrightarrow{P} 0$ and $1/\sqrt{m_n} \sum_{i=1}^{m_n} X_{ni} \xrightarrow{L} N_{0,1-\alpha}$.

In this instance, not only does $\bar{X}_n \xrightarrow{P} 0$, but even $1/\sqrt{k_n} \sum_{i=1}^{k_n} X_{ni} \xrightarrow{P} 0$, which is perhaps more than might be desired. Note that (4′) automatically prevails if the X_{ni} sum to C_n; in fact, $\operatorname{Cov}(X_{n1}, X_{n2}) = -[1/(k_n - 1)]$ in this case.

Define $Z_{ni} = X_{ni}/\sqrt{k_n}$. If (i) is replaced by (iii), $EX_{ni} = 0$, and (ii) still obtains, then $EZ_{ni} = 0$, $\sum_{i=1}^{k_n} \sigma_{Z_{ni}}^2 = 1$. Conditions (1) and (2) become

$$\max_{1 \leqq i \leqq k_n} |Z_{ni}| \xrightarrow{P} 0 \tag{1′}$$

and

$$\sum_{i=1}^{k_n} Z_{ni}^2 \xrightarrow{P} 1. \tag{2′}$$

Then, in view of theorems cited in Section 3, the conditions (2′) implies (1′) (and correspondingly (2) implies (1)) for infinitesimal row-wise *independent* r.v.'s, satisfying (ii) and (iii).

Of course, condition (i) precludes independence. Nonetheless, it should be verified for interchangeable r.v.'s satisfying (i) and (ii) that conditions (1) and (2) do not overlap. This may be seen from the following examples:

EXAMPLE 1. Let $(X_{n1}, X_{n2}, \cdots, X_{n,2n})$ be a random permutation of

$$(\sqrt{n}, -\sqrt{n}, 0, 0, \cdots, 0).$$

Then $\sum_{i=1}^{2n} X_{ni} = 0$, $1/2n \sum_{i=1}^{2n} X_{ni}^2 = 1$, but $\max_{1 \leqq i \leqq 2n} |X_{ni}|/\sqrt{2n} = 1/\sqrt{2}$.

EXAMPLE 2. Let $X = (X_{n1}, X_{n2}, \cdots, X_{n,2n}) = (0, 0, \cdots, 0)$ with probability $1 - p_n \to 1$, and otherwise let X be a random permutation of

$$\left(\frac{1}{\sqrt{p_n}}, -\frac{1}{\sqrt{p_n}}, \frac{1}{\sqrt{p_n}}, \cdots, -\frac{1}{\sqrt{p_n}}\right).$$

Then $\sum_{i=1}^{2n} X_{ni} = 0$, $E(X_{ni}^2) = 1$, and the X_{ni} are i.r.v.'s. Now

$$\max_{1 \le i \le 2n} \frac{|X_{ni}|}{\sqrt{2n}} = \frac{|X_{n1}|}{\sqrt{2n}} \xrightarrow{P} 0.$$

But $1/2n \sum_{i=1}^{2n} X_{ni}^2 = 0$ with probability $1 - p_n \to 1$ and hence converges to zero in probability.

5. Illustrations.

EXAMPLE 1. *Quantiles.* Let k, n be positive integers and $U_1, U_2, \cdots, U_{kn-1}$ independent r.v.'s each uniformly distributed on (0, 1). Take U_j^* to be the jth smallest of $(U_1, U_2, \cdots, U_{kn-1})$, $j = 1, 2, \cdots, kn - 1$. That is, $U_1^* \leqq U_2^* \leqq, \cdots, \leqq U_{kn-1}^*$ are the order statistics from a uniform or rectangular distribution. Designate the successive differences $U_i^* - U_{i-1}^*$ by V_i, $i = 1, 2, \cdots, kn$, where $U_0^* = 0$, $U_{kn}^* = 1$.

It is well known that $V_1, V_2, \cdots, V_{kn}$ are interchangeable random variables adding up to one. In fact, any $kn - 1$ of them have a joint density

$$\begin{aligned} f(v_1, v_2, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{kn}) &= (kn-1)! \quad \text{for} \quad \textstyle\sum_{j \ne i} v_j \leqq 1, v_j \geqq 0, \\ &= 0, \qquad \text{otherwise.} \end{aligned}$$

A routine but tedious calculation or a non-routine exciting application of the Poisson stochastic process yields

$$E[V_1^r] = \binom{kn-1+r}{r}^{-1}, \qquad r = 1, 2, \cdots,$$

$$E[V_1^2 V_2^2] = \frac{4(kn-1)!}{(kn+3)!},$$

$$E[V_1 V_2] = \frac{(kn-1)!}{(kn+1)!},$$

$$E[V_1^2 V_2] = \frac{2(kn-1)!}{(kn+2)!}.$$

Further, $V_1, \cdots, V_{kn}$ are i.r.v.'s and likewise $X_{n1}, \cdots, X_{n,k_n}$, where $k_n = kn$ and

$$X_{ni} = \frac{kn\left(V_i - \dfrac{1}{nk}\right)}{\sqrt{(kn-1)(kn+1)^{-1}}}, \qquad i = 1, 2, \cdots, k_n.$$

Moreover, $\sum_{i=1}^{k_n} X_{ni} = 0$ and $\sigma_{X_{ni}}^2 = 1$, $i = 1, \cdots, kn$. The prior array of expected values furnishes the estimates:

$$EX_{n1}^4 = O(n^4)E\left[V_1 - \frac{1}{kn}\right]^4 = O(n^4)O\left(\frac{1}{n^4}\right) = O(1)$$

and

$$\begin{aligned}
&\operatorname{Cov}(X_{n1}^2, X_{n2}^2)\\
&= O(n^4)\operatorname{Cov}\left[\left(V_1 - \frac{1}{kn}\right)^2, \left(V_2 - \frac{1}{kn}\right)^2\right]\\
&= O(n^4)\operatorname{Cov}\left[V_1^2 - \frac{2}{kn}V_1, V_2^2 - \frac{2}{kn}V_2\right]\\
&= O(n^4)\left\{\left[E(V_1^2V_2^2) - \frac{4}{kn}E(V_1^2V_2) + \frac{4}{k^2n^2}E(V_1V_2)\right] - \left[E(V_1^2) - \frac{2}{kn}E(V_1)\right]^2\right\}\\
&= O(n^4)O(n^{-6}) = O(n^{-2}).
\end{aligned}$$

If, now, $m_n = n$, it follows from Corollary 1 to Theorem 1 that

$$\frac{1}{\sqrt{n}}\frac{kn(U_n - 1/k)}{\sqrt{(kn-1)(kn+1)^{-1}}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{ni}$$

has a limiting normal distribution with mean zero and variance $1 - 1/k$. The same statement then applies to $k\sqrt{n}(U_n - 1/k)$.

Thus, the sample quantile U_n of order $1/k$ in a sample of $kn - 1$ from a rectangular distribution is asymptotically normal with expected value $1/k$ and variance $(k-1)/k^3n$.

Clearly, an analogous statement holds with $1/k$ replaced by any real number q in $(0, 1)$. This conclusion extends to other distributions than the rectangular, e.g., if the c.d.f. $F(x)$ has a continuous non-zero derivative at the unique solution x_k of $F(x) - 1/k$. These facts are, of course, well known.

Note, in addition, that

$$\begin{aligned}
EX_{n1}X_{n2} &= O(n^2)E[(V_1 - 1/kn)(V_2 - 1/kn)]\\
&= O(n^2)\left[\frac{1}{kn(kn+1)} - \frac{2}{kn}\frac{1}{kn} + \frac{1}{(kn)^2}\right]\\
&= O(n^2)O(n^{-3}) = o(1).
\end{aligned}$$

Thus, if (for specificity) $k = 2$, an application of Corollary 3 yields the conclusion that

$$\frac{1}{\sqrt{n}}\left[\frac{2n(U_n - \frac{1}{2})}{\sqrt{(2n-1)(2n+1)^{-1}}} - n\bar{X}_n\right] = \sqrt{n}\left[2\sqrt{\frac{2n+1}{2n-1}}(\tilde{X}_n - \tfrac{1}{2}) - \bar{X}_n\right]$$

is normally distributed in the limit with mean zero and variance $\frac{1}{2}$ where $\tilde{X}_n$ denotes the sample median. This appears to be new but hardly of overwhelming interest. A comparable result may be demonstrated in the case of a random casting of r_n objects into n cells referred to in Section 3.

EXAMPLE 2. *Ranks*. Let $R_1, \cdots, R_{k_n}$ be a random permutation of the integers $(1, 2, \cdots, k_n)$. Define

$$X_{ni} = \frac{R_i - \dfrac{k_n + 1}{2}}{\sqrt{\dfrac{k_n^2 - 1}{12}}}.$$

Then, $(R_1, \cdots, R_{k_n})$ and $(X_{n1}, \cdots, X_{n,k_n})$ each comprise a set of i.r.v.'s. Moreover,

$$\sum_{i=1}^{k_n} X_{ni} = 0, \qquad \sum_{i=1}^{k_n} X_{ni}^2 = 1,$$

and

$$\max_{1 \le i \le k_n} \frac{|X_{ni}|}{\sqrt{k_n}} \frac{\sqrt{12}}{\sqrt{k_n}} \frac{\left| k_n - \dfrac{k_n + 1}{2} \right|}{\sqrt{k_n^2 - 1}} = \frac{1}{\sqrt{k_n}} \sqrt{\frac{3(k_n - 1)}{(k_n + 1)}} \to 0.$$

A direct application of Corollary 2 of Theorem 1 yields the limiting normality (mean 0, variance $1 - \alpha$) of

$$\sum_{i=1}^{m_n} \frac{R_i - \dfrac{k_n + 1}{2}}{\sqrt{\dfrac{k_n^2 - 1}{12}}},$$

where $\lim_{n \to \infty} m_n/k_n = \alpha, 0 < \alpha < 1$, a familiar result.

REFERENCES

[1] E. SPARRE ANDERSEN, "On sums of symmetrically dependent random variables," *Skand. Aktuarietidskr.*, Vol. 36 (1953), pp. 123–138.
[2] E. SPARRE ANDERSEN, "On fluctuations of sums of random variables," I, II, *Math. Skand.*, Vol. 1 (1953), pp. 263–285, and Vol. 2 (1954) pp. 195–223.
[3] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[4] BRUNO DE FINETTI, "La previsions; ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, Vol. 7 (1937), pp. 1–68.
[5] GNEDENKO-KOLMOGOROFF, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley Publ. Co., Inc., Cambridge, Mass., 1954.
[6] MICHEL LOÈVE, *Probability Theory*, D. Van Nostrand, New York, 1955.
[7] R. VON MISES, "Uber Aufteilungs- und Besetzungswahrscheinlichkeiten," *Revue de la Faculte des Sciences de l'Universite d'Istanbul N.S.*, (1939), pp. 145–163.
[8] GOTTFRIED NOETHER, On a theorem by Wald and Wolfowitz, *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455–458.
[9] IRVING WEISS, "Limiting Distributions in Some Occupancy Problems," Technical Report No. 28, Applied Mathematics and Statistics Laboratory, Stanford University, Office of Naval Research Contract N6onr-25140, 1955.

LINEAR ESTIMATION FROM CENSORED DATA

By R. L. Plackett

University of Liverpool

1. Introduction. Suppose that a sample of n random variables is taken from a continuous probability distribution, whose density function is $f[(y - \mu)/\sigma]/\sigma$, where μ and σ are unknown. Arrange the variables in order of magnitude, and denote them by $y_1, y_2, \cdots, y_n$, where

$$y_1 < y_2 < \cdots < y_n.$$

We shall discuss the problem of estimating μ and σ from the k successive variables $y_u, y_{u+1}, \cdots, y_v$, where $v = u + k - 1$. This problem arises, for example, in life-testing, and some applications are described by Gupta [7].

When using the principal results derived here, the expected values of ordered variables are essential, but tables of these quantities for normal samples are, at present somewhat limited. However, recent studies by Berkson [1] have shown the importance of the logistic distribution, which closely resembles the normal, and some properties of ordered logistic variables are given in Section 2. We now turn to the main problem. If u and v are fixed, the best linear unbiased estimates of μ and σ can be calculated by least squares, given the expected value and dispersion matrix of the vector of ordered variables (Godwin [6], Lloyd [11], Gupta [7]. In general, special tables become necessary, and it seems desirable to obtain simple formulae when samples are moderate or large in size. This is achieved in Section 3, where asymptotic values of the coefficients of $y_u, y_{u+1}, \cdots, y_v$ are derived. An examination of the conditions involved is supplied in Section 4, by considering the limiting form of the maximum likelihood equations. Several illustrative numerical tables complete the paper.

2. Ordered logistic variables. The logistic distribution is defined by

$$L = \log\{p/(1 - p)\}, \tag{1}$$

where p is the probability of a value less than L. Suppose that $L(i; n)$ is the ith variable in ascending order in a sample of size n from this distribution. Then

$$\begin{aligned} \mathcal{E} \exp\{wL(i; n)\} &= \frac{n!}{(i-1)!(n-i)!} \int_0^1 \left(\frac{p}{1-p}\right)^w p^{i-1}(1-p)^{n-i}\, dp \\ &= \frac{(i-1+w)!(n-i-w)!}{(i-1)!(n-i)!}. \end{aligned} \tag{2}$$

Take logarithms, differentiate with respect to w, and put $w = 0$. The cumulants of $L(i; n)$ are

$$\kappa_j(i; n) = \frac{d^j}{dw^j} \log(i-1)! + (-1)^j \frac{d^j}{dw^j} \log(n-i)! \tag{3}$$

Received July 1, 1957; revised September 17, 1957.

and are thus expressible in terms of polygamma functions, tabulated for $j = 1, 2, 3, 4$ in [2]. For $(i - 1) > (n - i)$, we obtain

$$\kappa_1(i; n) = \frac{1}{(n - i + 1)} + \frac{1}{(n - i + 2)} + \cdots + \frac{1}{(i - 1)}, \tag{4}$$

$$\begin{aligned}\kappa_2(i; n) = \frac{\pi^2}{3} &- \left\{1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{(i - 1)^2}\right\} \\ &- \left\{1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{(n - i)^2}\right\},\end{aligned} \tag{5}$$

$$\kappa_3(i; n) = 2\left\{\frac{1}{(n - i + 1)^3} + \frac{1}{(n - i + 2)^3} + \cdots + \frac{1}{(i - 1)^3}\right\}, \tag{6}$$

$$\begin{aligned}\kappa_4(i; n) = \frac{2\pi^4}{15} &- 6\left\{1 + \frac{1}{2^4} + \frac{1}{3^4} + \cdots + \frac{1}{(i - 1)^4}\right\} \\ &- 6\left\{1 + \frac{1}{2^4} + \frac{1}{3^4} + \cdots + \frac{1}{(n - i)^4}\right\}.\end{aligned} \tag{7}$$

Suppose that $x(i; n)$ is the ith variable in ascending order in a sample of size n from the probability distribution whose density function is $f(x)$ and distribution function $F(x)$. Let α be fixed, $0 < \alpha < 1$, and define t by

$$\alpha = F(t). \tag{8}$$

We require the two following results. As $n \to \infty$, with $i = [n\alpha]$

$$\mathcal{E}x(i; n) = t + O(n^{-1}) \tag{9}$$

and

$$F\{\mathcal{E}x(i + 1; n)\} - F\{\mathcal{E}x(i; n)\} = 1/n + O(n^{-2}). \tag{10}$$

The proofs are based on the Taylor expansion of x, considered as a function of L, about the value $L = \kappa_1(i; n)$. This, after expectation, gives

$$\mathcal{E}x(i; n) = x^{(0)} + \tfrac{1}{2}x^{(2)}\kappa_2 + \tfrac{1}{6}x^{(3)}\kappa_3 + \tfrac{1}{24}x^{(4)}(\kappa_4 + 3\kappa_2^2) + \cdots, \tag{11}$$

where $x^{(j)}$ is the value at $L = \kappa_1(i; n)$ of the jth derivative of x with respect to L. Now

$$\kappa_1(i; n) = \tfrac{1}{2}\log\{(i - 1)i/(n - i)(n - i + 1)\} + O(n^{-2}) \tag{12}$$

whence

$$\kappa_1(i; n) = \lambda + O(n^{-1}), \tag{13}$$

where

$$\lambda = \log\{\alpha/(1 - \alpha)\}. \tag{14}$$

Also

$$\kappa_j(i; n) = O(n^{1-j}) \qquad (j = 2, 3, \cdots). \tag{15}$$

Assuming $x^{(2)}$ to be bounded, we can substitute (13) and (15) in (11) to obtain (9). As regards (10), we suppose that $x^{(2)}$ and $x^{(3)}$ are bounded, in which case

$$\mathcal{E}x(i+1;n) - \mathcal{E}x(i;n) = \left\{\frac{1}{i} + \frac{1}{(n-i)}\right\} x^{(1)} + O(n^{-2}). \tag{16}$$

On the further assumption that df/dx is bounded, (10) results.

We shall now consider the standard normal distribution in more detail. Denote its density function by $\phi(x)$ and distribution function by $\Phi(x)$. Here

$$x^{(1)} = \Phi(1-\Phi)/\phi, \tag{17}$$

$$x^{(2)} = x^{(1)}\{xx^{(1)} - (2\Phi - 1)\}, \tag{18}$$

$$x^{(3)} = (x^{(1)})^3 + 2xx^{(1)}x^{(2)} + x^{(2)}(1-2\Phi) - 2x^{(1)}\Phi(1-\Phi), \tag{19}$$

$$x^{(4)} = 5(x^{(1)})^2x^{(2)} + x^{(3)}\{2xx^{(1)} - (2\Phi - 1)\} + 2x^{(2)}\{xx^{(2)} - 2\Phi(1-\Phi)\} + 2x^{(1)}(2\Phi - 1)\Phi(1-\Phi). \tag{20}$$

These derivatives are all bounded, their maximum absolute values being given below.

$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
0.62666	0.07376	0.06724	0.04597

The absolute value of the remainder after $(j-1)$ terms of the series (11) is at most $\beta_j \max|x^{(j)}|/j!$, where β_j is the jth absolute moment about the mean of the ith ordered logistic variable in a sample of n. Since β_j is known when j is even and the inequality $(\beta_j)^{1/j} \leq (\beta_{j+1})^{1/(j+1)}$ is available when j is odd, we can thus assign bounds to $\mathcal{E}x(i;n)$ for all values of j. As an illustration, take $\mathcal{E}x(19;25)$.

j	Series (11) to j terms	Absolute maximum error
1	0.642835	0.007656
2	0.636781	0.002521
3	0.636656	0.000262

David and Johnson [5] express x as a function of Φ, and the value for $\mathcal{E}x(19;25)$ from the first four terms of the series on p. 236 of their paper is 0.636904. However, their formula is arranged as a power series in $(n+2)^{-1}$, and a similar rearrangement of (11) would be necessary before a full comparison of the two approaches can be made. This will be undertaken on another occasion.

3. Least squares estimation. Let t_i denote the expectation of $(y_i - \mu)/\sigma$. Write

$$f_i = f(t_i), \tag{21}$$

$$p_i = F(t_i), \tag{22}$$

and

$$q_i = 1 - p_i \qquad (i = u, u + 1, \cdots, v). \tag{23}$$

Let m be the vector of $(y_i - \mu)/\sigma$ for $i = u, u + 1, \cdots, v$ and put

$$t = \mathcal{E}m \tag{24}$$

and

$$V = \mathfrak{D}m = \mathcal{E}\{(m - \mathcal{E}m)(m - \mathcal{E}m)'\}. \tag{25}$$

In principle, t and V can be computed from the known function $f(x)$. The estimate of

$$\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix} \tag{26}$$

given by generalized least squares is

$$\theta^* = (A'V^{-1}A)^{-1}A'V^{-1}y, \tag{27}$$

where

$$A = [1 \quad t]. \tag{28}$$

As V is difficult to handle analytically, we replace it by W, a symmetric matrix whose elements are $\{a_i b_j\}$ for $i \leqq j$, where

$$a_i = p_i/f_i \qquad (i = u, u + 1, \cdots, v) \tag{29}$$

and

$$b_j = q_j/f_j \qquad (j = u, u + 1, \cdots, v). \tag{30}$$

Since $\mathfrak{D}y \sim W\sigma^2/n$, the unbiased estimate

$$\theta^+ = (A'W^{-1}A)^{-1}A'W^{-1}y \tag{31}$$

may be presumed to have the same asymptotic properties as θ^*. We therefore consider the limiting form of θ^+.

The inverse of W has been given by Hammersley and Morton [9]. Put

$$a_{u-1} = 0, \qquad a_{v+1} = 1, \qquad b_{u-1} = 1, \qquad b_{v+1} = 0, \tag{32}$$

Then

$$W^{-1} = \begin{bmatrix} c_u & d_u & 0 & 0 & \cdots & 0 \\ d_u & c_{u+1} & d_{u+1} & 0 & \cdots & 0 \\ 0 & d_{u+1} & c_{u+2} & d_{u+2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & & c_{v-1} & d_{v-1} \\ 0 & 0 & \cdots & & d_{v-1} & c_v \end{bmatrix}, \tag{33}$$

where

$$c_i = (a_{i+1}b_{i-1} - a_{i-1}b_{i+1}) \,/\, (a_i b_{i+1} - a_{i+1}b_i)(a_{i-1}b_i - a_i b_{i-1}) \tag{34}$$

and

$$d_i = 1/(a_i b_{i+1} - a_{i+1} b_i). \tag{35}$$

Denote $A'W^{-1}$ by G, and define

$$h_s = \tfrac{1}{2}(p_{s+1} - p_{s-1}) \qquad (s = u + 1, u + 2, \cdots, v - 1), \tag{36}$$

$$h_u = p_{u+1} - p_u , \tag{37}$$

and

$$h_v = p_v - p_{v-1} . \tag{38}$$

If the elements of G are considered as functions of p_u, p_{u+1}, $\cdots$, p_v then g_{1u} depends on p_u and p_{u+1}; g_{1s} on p_{s-1}, p_s, and p_{s+1}; and g_{1v} on p_{v-1} and p_v. In g_{1u}, put $p_{u+1} = p_u + h_u$; in g_{1s}, replace p_{s-1} by $p_s - h_s$ and p_{s+1} by $p_s + h_s$; and in g_{1v}, put $p_{v-1} = p_v - h_v$. The first and third substitutions are exact; the second one is approximate, but if $n \to \infty$ with $u = [n\alpha]$ and $v = [n\beta]$, the values of p_i tend to become equally spaced between α and β, as (10) shows. The elements in the second row of G are treated similarly, so that both elements in the ith column are now expressed as functions of p_i and h_i. Expanding by Taylor series as far as h_s^2 in the numerators of g_{1s} and g_{2s} and as far as h^2 elsewhere, the elements of G finally reduce, after a good deal of straightforward algebra, to the expressions given below. The primes signify differentiation with respect to p, so that

$$f' = \frac{d \log f}{dx} \tag{39}$$

and

$$ff'' = \frac{d^2 \log f}{dx^2}. \tag{40}$$

In calculating the elements of $A'W^{-1}A$, we pass from sums involving h to integrals involving dp.

4. Maximum likelihood estimation. The likelihood of y_u, y_{u+1}, $\cdots$, y_v is

$$\frac{n!}{(u-1)!(n-v)!}\left\{F\left(\frac{y_u-\mu}{\sigma}\right)\right\}^{u-1}\prod_{i=u}^{v}\frac{1}{\sigma}f\left(\frac{y_i-\mu}{\sigma}\right)\left\{1-F\left(\frac{y_v-\mu}{\sigma}\right)\right\}^{n-v}. \tag{41}$$

Denote by $\hat{\mu}$ and $\hat{\sigma}$ the maximum likelihood estimates of μ and σ, respectively. They satisfy the equations

$$-\frac{(u-1)\hat{\sigma}f\left(\dfrac{y_u-\hat{\mu}}{\hat{\sigma}}\right)}{nF\left(\dfrac{y_u-\hat{\mu}}{\hat{\sigma}}\right)} - \frac{\hat{\sigma}}{n}\sum_{i=u}^{v}\frac{d\log f}{dx}\left(\frac{y_i-\hat{\mu}}{\hat{\sigma}}\right) + \frac{(n-v)\hat{\sigma}f\left(\dfrac{y_v-\hat{\mu}}{\hat{\sigma}}\right)}{n\left\{1-F\left(\dfrac{y_v-\hat{\mu}}{\hat{\sigma}}\right)\right\}} = 0, \tag{42}$$

TABLE 1

Asymptotic value of $A' W^{-1}$

Row—Column	General Density	Normal Density
(1, u)	$f_u^2/p_u - f_u' f_u - \frac{1}{2} h_u f_u f_u''$	$f_u^2/p_u + t_u f_u + \frac{1}{2} h_u$
(1, s)	$-h_s f_s f_s''$	h_s
(1, v)	$f_v^2/q_v + f_v' f_v - \frac{1}{2} h_v f_v f_v''$	$f_v^2/q_v - t_v f_v + \frac{1}{2} h_v$
(2, u)	$t_u f_u^2/p_u - t_u f_u' f_u - f_u - \frac{1}{2} h_u (f_u' + f_u f_u'' t_u)$	$t_u f_u^2/p_u + t_u^2 f_u - f_u + h_u t_u$
(2, s)	$-h_s (f_s' + f_s f_s'' t_s)$	$2h_s t_s$
(2, v)	$t_v f_v^2/q_v + t_v f_v' f_v + f_v - \frac{1}{2} h_v (f_v' + f_v f_v'' t_v)$	$t_v f_v^2/q_v - t_v^2 f_v + f_v + h_v t_v$

TABLE 2

Asymptotic value of $A' W^{-1} A$

Row—Column	General Density	Normal Density
(1, 1)	$-\int_{p_u}^{p_v} ff''\, dp + f_v^2/q_v + f_v f_v' + f_u^2/p_u - f_u f_u'$	$f_v^2/q_v - t_v f_v + p_v + f_u^2/p_u + t_u f_u - p_u$
(1, 2) and (2, 1)	$-\int_{p_u}^{p_v} tff''\, dp + t_v f_v^2/q_v + t_v f_v f_v' + t_u f_u^2/p_u - t_u f_u f_u'$	$t_v f_v^2/q_v - t_v^2 f_v - f_v + t_u f_u^2/p_u + t_u^2 f_u + f_u$
(2,2)	$-\int_{p_u}^{p_v} t^2 ff''\, dp + t_v^2 f_v^2/q_v + t_v^2 f_v f_v' + p_v + t_u^2 f_u^2/p_u - t_u^2 f_u f_u' - p_u$	$t_v^2 f_v^2/q_v - t_v^3 f_v - t_v f_v + 2p_v + t_u^2 f_u^2/p_u + t_u^3 f_u + t_u f_u - 2p_u$

$$
(43) \qquad -\frac{(u-1)(y_u - \hat{\mu}) f\left(\dfrac{y_u - \hat{\mu}}{\hat{\sigma}}\right)}{nF\left(\dfrac{y_u - \hat{\mu}}{\hat{\sigma}}\right)} - \frac{k\hat{\sigma}}{n} - \frac{1}{n} \sum_{i=u}^{v} (y_i - \hat{\mu}) \frac{d \log f}{dx} \left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right) + \frac{(n-v)(y_v - \hat{\mu}) f\left(\dfrac{y_v - \hat{\mu}}{\hat{\sigma}}\right)}{n\left\{1 - F\left(\dfrac{y_v - \hat{\mu}}{\hat{\sigma}}\right)\right\}} = 0,
$$

where

$$
\frac{d \log f}{dx} \left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right)
$$

means the value at $(y_i - \hat{\mu})/\hat{\sigma}$ of the function $d \log f/dx$. The direct solution of (42) and (43) for normal samples has been described by Cohen [3], who used successive approximation; and, when $u = 1$, by Gupta [7], who calculated a special table which shortens the work. Halperin [8] has indicated conditions under which

(a) the maximum likelihood equations have a consistent set of solutions $\hat{\mu}$, $\hat{\sigma}$;

(b) $\sqrt{n}(\hat{\mu} - \mu)$ and $\sqrt{n}(\hat{\sigma} - \sigma)$ have a bivariate normal limit distribution;

(c) the dispersion matrix of the limit distribution is best in the sense of Cramér [4], §32.6.

The necessary assumptions involve derivatives of $f[(y - \mu)/\sigma]/\sigma$ with respect to μ and σ, and we shall suppose henceforth that they are satisfied.

We expand $(y_i - \hat{\mu})/\hat{\sigma}$ in a Taylor series about t_i and obtain

$$
\begin{aligned}
&- \frac{(u-1)\hat{\sigma}}{n}\left\{\frac{f_u}{p_u} + \left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)\left(\frac{f_u f_u'}{p_u} - \frac{f_u^2}{p_u^2}\right) + \frac{1}{2}\left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)^2 C\right\} \\
&- \frac{\hat{\sigma}}{n}\sum_{i=u}^{v}\left\{f_i' + \left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)f_i f_i'' + \frac{1}{2}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)^2 D_i\right\} \\
&+ \frac{(n-v)\hat{\sigma}}{n}\left\{\frac{f_v}{q_v} + \left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)\left(\frac{f_v f_v'}{q_v} + \frac{f_v^2}{q_v^2}\right) + \frac{1}{2}\left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)^2 E\right\} = 0,
\end{aligned}
\tag{44}
$$

$$
\begin{aligned}
&- \frac{(u-1)\hat{\sigma}}{n}\left\{\frac{t_u f_u}{p_u} + \left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)\left(\frac{f_u}{p_u} + \frac{t_u f_u f_u'}{p_u} - \frac{t_u f_u^2}{p_u^2}\right)\right. \\
&\left. + \frac{1}{2}\left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)^2 R\right\} - \frac{k\hat{\sigma}}{n} - \frac{\hat{\sigma}}{n}\sum_{i=u}^{v}\left\{t_i f_i'\right. \\
&\left. + \left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)(t_i f_i f_i'' + f_i') + \frac{1}{2}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)^2 S_i\right\} + \frac{(n-v)\hat{\sigma}}{n} \\
&\left\{\frac{t_v f_v}{q_v} + \left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)\left(\frac{f_v}{q_v} + \frac{t_v f_v f_v'}{q_v} + \frac{t_v f_v^2}{q_v^2}\right) + \frac{1}{2}\left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)^2 T\right\} = 0.
\end{aligned}
\tag{45}
$$

Here C, D_i, E, R, S_i and T are second-order derivatives with respect to x evaluated at points intermediate between $(y_i - \hat{\mu})\hat{\sigma}$ and t_i; and the primes have their previous meaning.

We assume that the second-order derivatives of

$$\frac{d \log f}{dx}, \quad x\frac{d \log f}{dx}, \quad \frac{f}{p}, \quad \frac{f}{q}, \quad \frac{xf}{p}, \quad \frac{xf}{q},$$

with respect to x, are functions of bounded variation. This condition is not satisfied if $F(x) = 0$ at a finite value of x, since then

$$\left|\frac{d^2}{dx^2}\left(\frac{f}{p}\right)\right| \to \infty$$

at the lower terminus of the distribution; nor if $F(x) = 1$ for finite x, since

$$\left|\frac{d^2}{dx^2}\left(\frac{f}{q}\right)\right| \to \infty$$

there. However, all is well for the normal and logistic distributions, as the following table shows.

Maximum absolute values of second-order derivatives

Distribution	$\frac{d \log f}{dx}$	$x \frac{d \log f}{dx}$	$\frac{f}{p}$	$\frac{f}{q}$	$\frac{xf}{p}$	$\frac{xf}{q}$
Normal	0	2.00	0.30	0.30	2.00	2.00
Logistic	0.19	1.00	0.10	0.10	0.50	0.50

Let α and β be fixed such that $0 < \alpha < \beta < 1$. We assume that $f(t) \geqq c > 0$ wherever $F^{-1}(\alpha) \leqq t \leqq F^{-1}(\beta)$. For any such t, let f, p and q be defined in accordance with (21), (22), and (23). Then

$$z^2 = pq/f^2 \tag{46}$$

has a finite maximum, which we denote by z_1^2. We proceed to derive the form taken by the maximum likelihood equations when $u = [n\alpha]$, $v = [n\beta]$, and $n \to \infty$.

Consider the variable

$$\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right) = \frac{(y_i - \mu - t_i\sigma) - (\hat{\mu} - \mu) - t_i(\hat{\sigma} - \sigma)}{\hat{\sigma}} \tag{47}$$

Given $\epsilon_1 > 0$, $\epsilon_2 > 0$, and ϵ_3 such that $\sigma > \epsilon_3 > 0$,

$$\begin{aligned} \Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| > \frac{\epsilon_1 + \epsilon_2}{\hat{\sigma} - \epsilon_3}\right\} &< \Pr\{|y_i - \mu - t_i\sigma| > \epsilon_1\} \\ &+ \Pr\{|(\hat{\mu} - \mu) + t_i(\hat{\sigma} - \sigma)| > \epsilon_2\} + \Pr\{|\hat{\sigma} - \sigma| > \epsilon_3\}. \end{aligned} \tag{48}$$

Typically, $i = [np]$, and as $n \to \infty$, $\frac{\sqrt{n}}{z}\left(\frac{y_i - \mu}{\sigma} - t\right)$ is asymptotically normal with zero mean and unit variance (Cramér [4], §28.5). According to (9), $(t_i - t)$ is $O(n^{-1})$ and so $\frac{\sqrt{n}}{z}\left(\frac{y_i - \mu}{\sigma} - t_i\right)$ has the same limit distribution. Hence

$$\Pr\{|y_i - \mu - t_i\sigma| > \epsilon_1\} \sim 2\Phi(-n^{1/2}\epsilon_1/\sigma z) \sim 2(\sigma z/n^{1/2}\epsilon_1)\phi(n^{1/2}\epsilon_1/\sigma z). \tag{49}$$

Similarly, by the asymptotic properties of $\hat{\mu}$ and $\hat{\sigma}$, there exist finite quantities z_2 and z_3 such that

$$\Pr\{|(\hat{\mu} - \mu) + t_i(\hat{\sigma} - \sigma)| > \epsilon_2\} \sim 2(\sigma z_2/n^{1/2}\epsilon_2)\phi(n^{1/2}\epsilon_2/\sigma z_2) \tag{50}$$

and

$$\Pr\{|\hat{\sigma} - \sigma| > \epsilon_3\} \sim 2(\sigma z_3/n^{1/2}\epsilon_3)\phi(n^{1/2}\epsilon_3/\sigma z_3). \tag{51}$$

Consequently, as $n \to \infty$,

$$\sum_{i=u}^{v} \Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| > \frac{\epsilon_1 + \epsilon_2}{\sigma - \epsilon_3}\right\} < 2(\beta - \alpha)n^{1/2}\sigma \sum_{j=1}^{3} (z_j/\epsilon_j)\phi(n^{1/2}\epsilon_j/\sigma z_j). \tag{52}$$

Thus, given $\epsilon > 0$ and $\delta > 0$, arbitrarily small, we can find N such that

$$\sum_{i=u}^{v} \Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| > \epsilon\right\} < \delta \qquad \text{for} \quad n \geqq N. \tag{53}$$

Therefore, by Boole's inequality,

$$\Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| \leqq \epsilon \quad \text{for} \quad i = u, u+1, \cdots, v\right\} \geqq 1 - \delta \quad \text{for} \quad n \geqq N. \tag{54}$$

When this event occurs, the terms in (44) and (45) which involve $\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)^2$ are negligible compared with the remaining terms, because C, D_i, E, R, S_i and T are all bounded. We therefore omit the squared terms, and replace sums by integrals as before (cf. Hoeffding [10]), at the same time making an Euler-Maclaurin adjustment on the coefficients of y_u and y_v, so as to correct for bias in the estimates and bring the results into line with those previously obtained. The linearized form taken asymptotically by the maximum likelihood equations is then given by the coefficients in Tables 1 and 2 on replacing h_i by $1/n$ throughout. We shall use μ^0 and σ^0 to denote the corresponding linearized estimates. The asymptotic dispersion matrix of the maximum likelihood estimates is σ^2/n times the inverse of the equations for μ^0 and σ^0.

These results show, not only that the maximum likelihood estimates of μ and σ are asymptotically linear, but also that the best linear unbiased estimates

TABLE 3A

Coefficients of ordered variables when estimating the mean. θ^* above, θ° below

k	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
2	−1.8634	2.8634								
	−2.1547	3.0554								
3	−0.6596	−0.2138	1.8734							
	−0.7487	−0.2248	1.9309							
4	−0.2923	−0.0709	0.0305	1.3327						
	−0.3304	−0.0780	0.0362	1.3543						
5	−0.1240	−0.0016	0.0549	0.0990	0.9718					
	−0.1418	−0.0055	0.0567	0.1071	0.9797					
6	−0.0316	0.0383	0.0707	0.0962	0.1185	0.7078				
	−0.0394	0.0366	0.0718	0.1003	0.1261	0.7100				
7	0.0244	0.0636	0.0818	0.0962	0.1089	0.1207	0.5045			
	0.0222	0.0633	0.0829	0.0988	0.1131	0.1270	0.5045			
8	0.0605	0.0804	0.0898	0.0972	0.1037	0.1099	0.1161	0.3424		
	0.0616	0.0812	0.0911	0.0992	0.1066	0.1137	0.1210	0.3423		
9	0.0843	0.0921	0.0957	0.0986	0.1011	0.1036	0.1060	0.1085	0.2101	
	0.0877	0.0934	0.0973	0.1004	0.1033	0.1060	0.1089	0.1120	0.2113	
10	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
	0.1048	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1048

are asymptotically normal and efficient. In order to compute the coefficients of the ordered variables, only tables of $f(x)$, $F(x)$ and t_i are necessary. For normal samples, Teichroew [14] gives t_i to 10 D for $n \leqq 20$, and an extension to $n \leqq 100$ with 24 D is being prepared (Ruben [12]). For logistic samples, explicit formulae have already been given.

5. Numerical tables. Tables 3A and 3B refer to the estimation of θ from the smallest k observations in a sample of size $n = 10$ from a normal distribution. They give the coefficients of $y_1, y_2, \cdots, y_k$ for $k \leqq 10$ in

(i) the best linear unbiased estimate, $\theta^* = \begin{bmatrix} \mu^* \\ \sigma^* \end{bmatrix}$,

(ii) the linearized maximum likelihood estimate, $\theta^0 = \begin{bmatrix} \mu^\circ \\ \sigma^\circ \end{bmatrix}$

Table 4 gives the coefficients of μ and σ in the expectation of θ^0. Suppose in general that

$$\mathcal{E}\theta^0 = B\theta \tag{55}$$

Then $B^{-1}\theta^0$ is an unbiased estimate of θ, and the efficiencies of its elements, relative to μ^* and σ^* respectively, have been calculated from the table of $\mathfrak{D}m$ in Sarhan and Greenberg [13] when, as above, $n = 10$, $u = 1$, and $v = 2, 3, \cdots 10$. These efficiencies never fall below 0.9998, a result which suggests that θ^0, corrected for bias, can be used in place of θ^*, with negligible loss of efficiency, for all sample sizes of practical importance.

TABLE 3B

Coefficients of ordered variables when estimating the standard deviation. θ^* above, θ° below

k	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
2	−1.8608	1.8608								
	−2.1366	2.0404								
3	−0.9625	−0.4357	1.3981							
	−1.0767	−0.4586	1.4738							
4	−0.6520	−0.3150	−0.1593	1.1263						
	−0.7190	−0.3330	−0.1611	1.1681						
5	−0.4419	−0.2491	−0.1362	−0.0472	0.9243					
	−0.5374	−0.2631	−0.1414	−0.0425	0.9499					
6	−0.3931	−0.2063	−0.1192	−0.0501	0.0111	0.7576				
	−0.4266	−0.2175	−0.1250	−0.0498	0.0180	0.7740				
7	−0.3252	−0.1758	−0.1058	−0.0502	−0.0006	0.0469	0.6107			
	−0.3513	−0.1849	−0.1114	−0.0517	0.0022	0.0545	0.6218			
8	−0.2753	−0.1523	−0.0947	−0.0488	−0.0077	0.0319	0.0722	0.4746		
	−0.2963	−0.1600	−0.0998	−0.0510	−0.0069	0.0358	0.0799	0.4830		
9	−0.2364	−0.1334	−0.0851	−0.0465	−0.0119	0.0215	0.0559	0.0936	0.3423	
	−0.2539	−0.1399	−0.0897	−0.0490	−0.0122	0.0234	0.0602	0.1009	0.3505	
10	−0.2044	−0.1172	−0.0763	−0.0436	−0.0142	0.0142	0.0436	0.0763	0.1172	0.2044
	−0.2196	−0.1231	−0.0807	−0.0462	−0.0151	0.0151	0.0462	0.0807	0.1231	0.2196

TABLE 4

Expectation of θ^0

k	μ^0		σ^0	
	μ	σ	μ	σ
2	0.9007	0.2560	−0.0962	1.2446
3	0.9574	0.1104	−0.0616	1.1492
4	0.9821	0.0539	−0.0450	1.1066
5	0.9962	0.0260	−0.0346	1.0827
6	1.0054	0.0108	−0.0270	1.0678
7	1.0119	0.0022	−0.0208	1.0583
8	1.0166	−0.0023	−0.0153	1.0529
9	1.0204	−0.0038	−0.0097	1.0523
10	1.0243	0.0000	0.0000	1.0668

TABLE 5

$1/n$	h_1	h_2	h_3	h_4	h_5
0.5000	0.4274				
0.3333	0.3013	0.3013			
0.2500	0.2316	0.2326			
0.2000	0.1879	0.1888	0.1897		
0.1667	0.1580	0.1588	0.1597		
0.1429	0.1362	0.1370	0.1378	0.1378	
0.1250	0.1198	0.1204	0.1212	0.1212	
0.1111	0.1068	0.1074	0.1081	0.1082	0.1082
0.1000	0.0964	0.0970	0.0976	0.0976	0.0976

Table 5 also refers to normal samples. Used in conjunction with the relation

$$h_i = h_{n+1-i}, \tag{56}$$

it gives the values of h_i for $n = 2, 3, \cdots 10$ and $1 \leqq i \leqq n$. That there is close agreement between θ^+ and θ^0 can be inferred from Table 5 in particular and (10) in general.

6. Acknowledgements. I am grateful to Mr. C. J. Taylor for doing all the calculations; and to the referee for a correction to my maximal remainder in (11), and other helpful remarks.

REFERENCES

[1] J. Berkson, "Why I prefer logits to probits," *Biometrics*, Vol. 7 (1951), pp. 327–339.
[2] *British Association Mathematical Tables*, Vol. 1 (1931).
[3] A. C. Cohen, Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 557–569.
[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[5] F. N. David and N. L. Johnson, "Statistical treatment of censored data. Part 1. Fundamental formulae," *Biometrika*, Vol. 41 (1954), pp. 228–240.

[6] H. J. Godwin, "On the estimation of dispersion by linear systematic statistics," *Biometrika*, Vol. 36 (1949), pp. 92–100.
[7] A. K. Gupta, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 260–273.
[8] M. Halperin, "Maximum likelihood estimation in truncated samples," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 226–238.
[9] J. M. Hammersley and K. W. Morton, "The estimation of location and scale parameters from grouped data," *Biometrika*, Vol. 41 (1954), pp. 296–301.
[10] W. Hoeffding, "On the distribution of the expected values of the order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.
[11] E. H. Lloyd, "Least squares estimation of location and scale parameters using order statistics," *Biometrika*, Vol. 39 (1952), pp. 88–95.
[12] H. Ruben, Personal communication.
[13] A. E. Sarhan and B. G. Greenberg, "Estimation of location and scale parameters from singly and doubly truncated samples. Part I. The normal distribution up to samples of size 10," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 427–451.
[14] D. Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.

SEMIMARTINGALES OF MARKOV CHAINS

By John G. Kemeny and J. Laurie Snell[1]

Dartmouth College

1. Introduction. We shall deal throughout this paper with absorbing Markov chains with a finite number of states. An absorbing Markov chain is one that has a set of "boundary" states which once reached cannot be left, and such that from any state the process reaches the boundary with probability 1. The chain is given by the transition matrix P, with entries p_{ij}.

More precisely, a state i is a "boundary" state if $p_{ii} = 1$. The remaining states will be called "interior" states. We must require that it is possible to reach the boundary from every interior state, not necessarily in one step. We assume, that there are r absorbing states and s interior states. The set of boundary states will be called B, the set of interior states I.

An upper semimartingale is a function on the states of the chain, such that the expected value of the function after one step from any state is greater than or equal to the value of the function at the state. A lower semimartingale is defined similarly, with the inequalities reversed. A martingale is a function on the states that is both an upper and a lower semimartingale.

A function on the states can be conveniently represented by a column vector. Such a vector z is an *upper semimartingale* if $Pz \geqq z$, a *lower semimartingale* if $Pz \leqq z$, and a *martingale* if $Pz = z$.

We assume that a set of nonnegative *boundary values* is assigned to the elements of B, v_j being assigned to state j. We denote by U the set of all nonnegative upper semimartingales and by U^* the set of all nonnegative lower semimartingales having the right boundary values. Thus U is the set of all vectors such that

$$\text{(a) } Pz \geqq z, \quad \text{(b) } z \geqq 0, \quad \text{(c) } \{z\}_j = v_j \quad \text{for} \quad j \in B.$$

The set U^* consists of the vectors satisfying conditions (b) and (c), and condition (a) with the inequality sign reversed.

Throughout the paper $\{z\}_j$ will denote the jth componnet of the vector z. Inequality signs between vectors will assert that the inequality holds componentwise.

A representation theorem will be developed for all nonnegative semimartingales with the prescribed boundary values in terms of martingales of modified chains. A modified chain is one obtained by adding interior states to the boundary, and assigning value 0 to them. The representation is unique and leads to a simple geometric interpretation. U will be represented (except in certain de-

Received July 24, 1957.

[1] This research was supported by the National Science Foundation through a grant given to the Dartmouth Mathematics Project.

generate cases) by a convex cubic s dimensional polyhedron. In degenerate cases the polyhedron reduces to smaller s dimensional polyhedra, including an s-simplex in the most degenerate case. U^* will be obtained from a reflection of U through the unique martingale.

These results will be applied to a treatment of certain sequential games, and to discrete subharmonic functions. In the latter application we will see that discrete subharmonic functions can be expressed as convex combinations of certain harmonic functions. And it is well known that the discrete harmonic function for given boundary values may be interpreted as the expected final value of a random walk. Hence we have a method of obtaining all discrete subharmonic functions in terms of certain random walks.

2. The basic semimartingales. Let T be a subset of I and denote by $P(T)$ the transition matrix obtained from P by changing the states in T into absorbing states. Let $Q(T) = \lim_{n\to\infty} [P(T)]^n$. Then the ijth entry of $Q(T)$, $q_{ij}(T)$ represents the probability, that starting at state i, the process will reach state j before reaching any element of T. Let $q_j(T)$ denote the jth column of $Q(T)$. Then since

$$Q(T) = P(T)\cdot Q(T),$$

$$q_{ij}(T) = \begin{cases} \sum_k p_{ik} q_{kj}(T), & i \not\in T, \\ 0, & i \in T, \end{cases}$$

we see that

$$Pq_j(T) \geqq q_j(T), \qquad j \in B.$$

Thus $q_j(T)$ is an upper semimartingale. It has the boundary value of 0 on all states of B except j, and has the value of 1 on this state. Thus the vector $z(T)$ given by

$$z(T) = \sum_{j=1}^{r} v_j\, q_j(T)$$

is a nonnegative upper semimartingale with the prescribed boundary values; *i.e.*, for each T, $z(T)$ is an element of U. We shall refer to $z(T)$ as a *basic upper semimartingale*.

The vector $z(T)$ may be interpreted in a game played as follows: The process starts in a given state, and continues until it reaches a state in T, or a state in B, and is then stopped. If it stops at a state j in B the player receives v_j; if it stops at a state in T, he receives 0. Then $\{z(T)\}_i$ represents the expected value of the game to the player starting at state i. We shall appeal to this interpretation for certain simple results, rather than give detailed proofs. For example:

LEMMA 1. *Assume that T_1 and T_2 are subsets of I such that $T_1 \subseteq T_2$. Then $z(T_1) \geqq z(T_2)$.*

From this interpretation we can easily determine $z(\phi)$ and $z(I)$. If $T = \phi$, then our game is always played till the boundary is reached, and hence $z(\phi)$ is the unique martingale with the prescribed boundary values. If $T = I$, then we can never reach the boundary from I. Hence

$$\{z(I)\}_i = \begin{Bmatrix} v_i, & i \in B \\ 0, & i \in I \end{Bmatrix}.$$

It can be seen from Lemma 1 that $z(\phi)$ and $z(I)$ are the largest and smallest $z(T)$, respectively. Since we will see later that all elements of U are convex combinations of $z(T)$'s, we see that $z(\phi)$ is the maximal and $z(I)$ the minimal element of U.

3. A special case. We shall first solve the problem of describing U for the case where the following hypothesis is satisfied.

HYPOTHESIS **A**: *The boundary values v_j are all positive, and for any state i in I there is at least one j in B such that $p_{ij} > 0$.*

In the case that hypothesis **A** is satisfied the game interpretation for $z(T)$ makes it clear that the following lemma holds.

LEMMA 2. *Under hypothesis* **A**, *the $z(T)$ have the property that*

$$\{Pz(T)\}_i = \{z(T)\}_i > 0 \qquad \textit{for } i \in I - T$$

and

$$\{Pz(T)\}_i > \{z(T)\}_i = 0 \qquad \textit{for } i \in T.$$

Thus for each component of $z(T)$ exactly one of the equalities in the defining conditions (a) and (b) of U holds.

LEMMA 3. *Let $x_1, x_2, \cdots, x_n$ be distinct nonnegative vectors. Let W_i be the set of components of x_i which are 0. Assume that if $W_i \subseteq W_k$ then $x_i \geqq x_k$. If so the vectors are convexly independent.*

PROOF. Assume that $x_i = \sum_k a_k x_k$ with $a_k > 0$ and $k \neq i$ and $\sum_k a_k = 1$. Then a component of x_i can be 0 only if all the x_k have this component 0. Hence $W_i \subseteq W_k$, and $x_i \geqq x_k$ for all k. But this can only be true if $x_i = x_k$ for all k, contrary to hypothesis.

DEFINITION. A convex n-dimensional polyhedron is *cubic* if in every j dimensional face for each $j - 1$ dimensional subface there is a unique nonintersecting $j - 1$ dimensional subface ($j = 1, 2, \cdots, n$).

THEOREM 1. *If hypothesis A is satisfied, then U is a convex cubic polyhedron with 2^r corner points. These corner points are the $z(T)$ for $T \subseteq I$.*

PROOF. We observe first that the 2^r $z(T)$ are distinct and convexly independent. This follows from Lemmas 1 and 3. We shall now prove that the convex set spanned by the $z(T)$ is a cubic polyhderon.

A j dimensional face of the convex set spanned by the $z(T)$ is determined by picking any $r - j$ interior states and requiring that one of the equalities

(a) $\{Pz(T)\}_i = \{z(T)\}_i$,

(b) $\{z(T)\}_i = 0$

hold for each i in the set chosen. To obtain a $j - 1$ dimensional subface of this face we impose an equality on one more component—say k. It follows from hypothesis **A** that $Pz > 0$. Hence it is not possible to have equality (a) and (b) for the same state. Hence this $j - 1$ dimensional face cannot intersect the face obtained by choosing the other equality for the kth component. By Lemma 2 we can find a $z(T)$ which has any prescribed set of equalities one for each of the pairs (a) and (b). Thus any $j - 1$ dimensional face obtained by choosing an equality for a component $i \neq k$ must intersect that obtained by choosing an equality for component i. Thus the set spanned by $z(T)$ satisfies the conditions for a cubic polyhedron.

To complete the proof of the theorem we must show that if z is in U then it must be in the cubic polyhedron spanned by the basic upper semimartingales $z(T)$. But if z is in U it must satisfy

$$\text{(a) } Pz \geqq z,$$

$$\text{(b) } z \geqq 0.$$

Thus for each interior state i it must lie between the hyperplane obtained by requiring $\{Pz\}_i = \{z\}_i$ and the hyperplane obtained by requiring $\{z\}_i = 0$ But this means that z must lie between each pair of opposite faces in the cubic polyhedron spanned by $z(T)$. Hence it must lie inside of this polyhedron.

Definition. A sequence $T_0 \subset T_1 \subset \cdots \subset T_k$ of subsets of I is called a *chain*. The corresponding sequence of corner points $z(T_0), z(T_1), \cdots, z(T_k)$ is called a z-chain. If $k = s$, the chain is called *maximal*.

It is clear that the elements of a z-chain are linearly independent and hence span a simplex. A simplex spanned by a z-chain will be called a *z-simplex*.

Lemma 4. *Every face (of every dimension) of the cube U has a maximal element.*

Proof. In the s-dimensional cube U, every j-face ($j = 0, 1, \cdots, s$) is a j dimensional cube. This is clear from the definition of the cubic polyhedron. The face of the cube is specified by imposing equalities of type (a) or (b) on $r - j$ components.

Since we have a polyhedral set, it suffices to show that there is a maximal corner. The corners are specified by imposing equalities of one of the two types on each of the j remaining components. It is a direct consequence of Lemma 1 that a corner $z(T)$ is maximal if its T is minimal. Hence we get a maximal corner by imposing equalities (a) on all j of the remaining components.

Lemma 5. *The intersection of two z-simplexes (if not empty) is a z-simplex which is a common face of the two original simplexes.*

Proof. Let $T_0 \subset T_1 \subset \cdots \subset T_k$ and $T_0' \subset T_1' \subset \cdots \subset T_{k'}'$. Let the two simplexes be determined by the corresponding z's. If there is a nonempty set of T's that the two chains have in common, then they span a common face. We will show that this is the intersection of the two simplexes.

It will suffice to show that all the remaining corners of the second simplex (if any) lie outside the first simplex. Let T' be one of the sets in the second chain that is not in the first chain. If $z(T')$ lies in the first simplex, then it is a convex combination of its corners. But this is impossible, since the $z(T)$'s are convexly independent. This completes the proof.

LEMMA 6. *Every point of U lies in at least one z-simplex.*

PROOF. Let z_0 be a point of U. Starting with ϕ we will construct a chain so that z_0 will lie in the simplex spanned by the corresponding z-chain.

First of all, draw a line from $z(\phi)$ through z_0 and continue it till it hits a face of U (of dimension less than s). Say it meets this face in the point z_1. Then z_0 is in the set spanned by $z(\phi)$ and z_1. In this face we pick the maximal point $z(T_1)$, which exists by Lemma 4, and draw a line from it through z_1 till we hit a face of lower dimension at a point z_2. Since z_1 lies in the set spanned by $z(T_1)$ and z_2, we know that z_0 is in the set spanned by $z(\phi)$ and $z(T_1)$ and z_2. We iterate this procedure until some z_n turns out to be a corner $z(T_n)$. This must happen, since the dimension of the face decreases at each step. Then we will have z_0 in the set spanned by $z(\phi), z(T_1), \cdots, z(T_n)$.

At each step we first introduced the minimal T in the face, hence the T's are monotone decreasing and hence form a chain. Thus the corners we found form a z-chain and the set they span is a z-simplex, which contains z_0.

THEOREM 2. *Any z_0 in U can be written uniquely as*

$$z_0 = \sum_{j=0}^{k} a_j\, z(T_j),$$

with $a_j > 0$ and $\sum a_j = 1$, where the $z(T)$'s used form a z-chain.

PROOF. Let z_0 be any point in U. By Lemma 6 it lies in at least one z-simplex Form the intersection of all z-simplexes that contain z_0. This intersection is not empty and hence by Lemma 5 it is a common face of all the z-simplexes. This smallest possible z-simplex serves the purpose of our representation. Its corners form a z-chain, and we can write z_0 as a convex combination of these. The weights a_j must all be positive, or else the point z_0 would lie in a smaller z-simplex.

To show the uniqueness of our representation we need only recall that the representation of a point in a simplex in our (barycentric) representation is unique. To get a representation of our form, the z-chain used must span a simplex containing z_0. Hence the minimal simplex is a face of it. Hence the a_j's can be all positive only if the simplex is the minimal one we found. This establishes the unique representation.

It is worth remarking that the theorem established only the uniqueness of the smallest z-simplex containing z_0. If this simplex is of a dimension smaller than s, then it is a common face of several z-simplexes. If hypothesis **A** is satisfied, then there are $s!$ maximal z-chains, and correspondingly $s!$ maximal z-simplexes. The cube is divided into these, and they overlap only in that they have common faces of lower dimension. If a point is in the interior of one of the maximal simplexes, then it is expressed by putting positive weights on all $s + 1$ corners. If it is on a face, we must apply the same consideration to the smaller simplexes in which it lies.

4. The general case. If we drop hypothesis **A**, most of the previous considerations still apply. However, the argument as to the distinctness of the $z(T)$'s breaks down. But by a continuity argument we can see that any case where the hypothesis is not fulfilled is a limiting case of ones where the hypothesis holds, and hence the worst that can happen is that some of the $z(T)$'s coincide, and hence we have fewer corners on U. While it is still a polyhedron of dimension s, it need not be cubic, and there will be fewer distinct z-simplexes. We will show how the distinct z-simplexes can be found in the general case.

Let B^* be the set of boundary points which have nonzero values assigned. We assume that B^* is not empty.

DEFINITION. A set T is *fundamental* if from any point in $I - T$ it is possible to reach B^* without going through T.

Let T be any set which is not fundamental. Add to T all states which are cut off from the set B^* by T. The new set T' is fundamental and $z(T)$ and $z(T')$ are the same. On the other hand the $z(T)$'s whose T is fundamental have 0 components exactly on T, and hence are distinct. Thus the extreme points of U are given by the $z(T)$'s with T fundamental.

LEMMA 7. *There exists at least one z-simplex of dimension s.*

PROOF. Let the index of an interior state be the minimum number of steps required to reach a state of B^* from it. Reorder the states in such a way that their indices are nonincreasing. Then from any state it must be possible to reach the boundary without going through a state appearing earlier in the sequence. Let T_j be the set of the first j states. Then $T_0, T_1, T_2, \cdots, T_s$ is a complete chain with all the T_j's fundamental. Hence $z(T_0)$, $z(T_1)$, $\cdots$, $z(T_s)$ form the corner points of a z-simplex of dimension s.

Lemma 7 is all that is needed to insure the construction used in Theorem 2. Hence the representation theorem applies equally well to the general case. The lemma also establishes that even in the degenerate cases U has dimension s.

5. The set of lower semimartingales U^*. The set U^* of all lower semimartingales having prescribed boundary conditions is determined by replacing the condition

$$\text{(a) } Pz \geqq z$$

by the condition

$$\text{(a}') \ Pz \leqq z.$$

It is easy to determine the set U^* from what we know about U. Each face (of dimension $j - 1$) of U lies in a hyperplane determined by an equality $\{Pz\}_i = \{z\}_i$ or $\{z\}_i = 0$. The latter type faces lie in the coordinate planes. The former normally protrude, and they have the martingale $z(\phi)$ as maximal corner. U^* is obtained by taking the set that lies on the other side of the hyperplanes $\{Pz\}_i = \{z\}_i$. This is an s dimensional cone with $z(\phi)$ as minimal element. Thus U^* is the reflection of U through $z(\phi)$—and its linear extension. We also see that the martingale is the maximal upper semimartingale and the minimal lower semimartingale—the only point U and U^* have in common. It is possible to represent these lower semimartingales in terms of the $z(T)$'s. In fact let z^* be any point in U^*. Then a line from z^* through $z(\phi)$ will intersect a coordinate plane in a point z_1 in U. Then z may be uniquely written in the form

$$z^* = z(\phi) + A(z(\phi) - z_1),$$

where A is a nonnegative constant. On the other hand, by Theorem 2 $z_1 = \sum_j a_j z(T_j)$, $a_j > 0$, and $\sum a_j = 1$. Thus

$$z^* = z(\phi) + A(z(\phi) - \sum a_j z(T_j)),$$

and A and the a_k's are unique.

We can summarize this by saying that we have a unique representation for lower semimartingales:

$$z^* = \sum_{j=0}^{k} a_j z(T_j),$$

where $a_j < 0$ for $j \neq 0$, and $\sum a_j = 1$, with $\phi = T_0 \subset T_1 \subset \cdots \subset T_k$ forming a chain.

6. Arbitrary boundary values. We have assumed that specific boundary values were given. The particular convex polyhedron obtained for U depends on these boundary values. However, the extreme points $z(T)$ are easily obtained from $Q(T)$ for any choice of boundary values. In fact $z(I)$ is the vector with r components given by the boundary values and 0 for all other components. The vectors $z(T)$ are given by $z(T) = Q(T)z(I)$. The matrix $Q(T)$ does not depend upon the boundary values, thus when we find these $Q(T)$'s we have essentially solved the problem for all possible conditions.

7. Two examples. We shall give here two examples, one where hypothesis **A** is satisfied and one where it is not. For the first case let P be

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

The states $B = \{1, 2\}$ are the boundary states and $\mathrm{I} = \{3, 4\}$ are the interior states. Assume that $v_1 = 2$ and $v_2 = 1$. The corner points are given by

$$\text{Martingale} = z(\phi) = \begin{pmatrix} 2 \\ 1 \\ \frac{7}{5} \\ \frac{9}{5} \end{pmatrix}; \quad z(\{3\}) = \begin{pmatrix} 2 \\ 1 \\ 0 \\ \frac{4}{3} \end{pmatrix}; \quad z(\{4\}) = \begin{pmatrix} 2 \\ 1 \\ \frac{1}{2} \\ 0 \end{pmatrix}; \quad z(\{3, 4\}) = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

The set of all upper semimartingales consists of the set of all vectors

$$z = \begin{pmatrix} 2 \\ 1 \\ x \\ y \end{pmatrix},$$

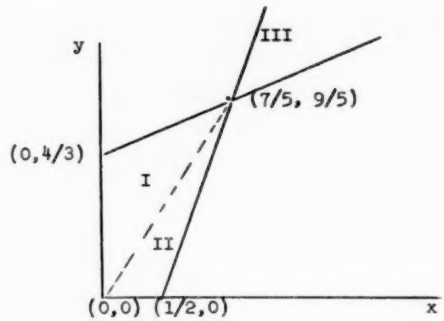


FIG. 1

where (x, y) is a point in the quadrilateral in Fig. 1. There are two maximal chains $\{0\}, \{3\}, \{3, 4\}$ and $\{0\}, \{4\}, \{3, 4\}$.

The regions above and below the dotted line, indicated by I and II, respectively, are the corresponding simplexes. The lower semimartingales are given by region III.

As an example of a case where we do not get a cubic polyhedron we consider the problem of random walk on the line with states 0 and $s + 1$ absorbing. Then the interior states are $I = (1, 2, \cdots, s)$. We require that $v(0) = 0$ and $v(s + 1) = 1$. It is clear that many subsets of I are not fundamental. In fact the only fundamental sets are the sets ϕ and $T_j = \{1, 2, \cdots, j\}$ for $1 \leqq j \leqq s$. Thus U is the s-dimensional simplex with corners $z(\phi), z(T_1), z(T_2), \cdots, z(T_s)$ These corner points are easily found from the ruin probabilities. They have coordinates for the interior states given by

$$\{z(T_j)\}_i = \begin{cases} 0, & i \leqq j, \\ \dfrac{i - j}{s + 1 - j} & j < i. \end{cases}$$

Thus any upper semimartingale vector

$$z = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_s \\ a_{s+1} \end{pmatrix},$$

with $a_0 = 0, a_{s+1} = 1$, is given by $z = \sum_{j=0}^{s} t_j z(T_j), \sum t_j = 1$. In this case it is easy to reverse the process and to find the t's from the z's. In fact for given z

$$t_0 = (s + 1)a_1,$$

$$t_j = (s + 1 - j)[a_{j+1} - 2a_j + a_{j-1}], \qquad 1 \leqq j \leqq s.$$

8. Application to sequential games. Consider an absorbing chain with r absorbing states and s interior states. Assume that we are given a vector

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{r+s} \end{pmatrix}$$

which determines the following game: The player starts in one of the states of the chain. If he is at an interior state i, he may either quit and collect v_i, or he may move on with the given transition probabilities. If he reaches a boundary state i, he collects v_i and the game ends. Let z_i be the value of the game to the player if he starts in state i. We wish to find the vector

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{r+s} \end{pmatrix}.$$

This is a special case of a problem considered in [2]. However, we can give a more precise description of the solution in the case considered here. It is clear that

$$z = \max [v, Pz], \tag{1}$$

since the player may by quitting or continuing have either of these. We shall now find a z having this property and then show that it is unique. Define

$$\bar{z} = \inf_z (z \geqq v, z \geqq Pz). \tag{2}$$

That is, $\bar{z}$ is the smallest lower semimartingale greater than v. If $\bar{z}$ did not have the property (1), then we could obtain a smaller semimartingale greater than v by replacing $\{\bar{z}\}_i$ by max $(v_i, \{P\bar{z}\}_i)$ in any component i for which $\{\bar{z}\}_i >$ max $(v_i, \{P\bar{z}\}_i$. Hence $\bar{z}$ must have the property (1).

Assume now that, for some other z, (1) is true. Let T be the set of interior states for which $\{z\}_i = v_i$, then

$$P(T)z = z$$

and thus

$$Q(T)z = \lim_{n \to \infty} [P(T)]^n z = z.$$

On the other hand,

$$P(T)\bar{z} \leqq \bar{z}$$

so that

$$Q(T)\bar{z} \leqq \bar{z}.$$

From the interpretation of $Q(T)$ (see Sec. 2), we know that $Q(T)z$ depends only on the components of z in $B \cup T$. And this is the set where $z = v$. Hence,

$$Q(T)z = Q(T)v.$$

Thus

$$z = Q(T)z = Q(T)v \leqq Q(T)\bar{z} \leqq \bar{z}.$$

But since $z \geqq v$ and $\geqq Pz$ we see from (2) that $z \geqq \bar{z}$. Therefore, $z = \bar{z}$. Hence $\bar{z}$ is the unique vector satisfying (1), and its components are the value of the game for various starting positions.

The optimal strategy is to continue on any component where $v_j < \{\bar{z}\}_j$.

A similar analysis shows that if the player wishes to minimize his fortune he should find the largest upper semimartingale z less than or equal to v and play only on states i such that $\{z\}_i < v_i$. This latter problem has application to statistical decision theory (see [2]).

For the first example given in Sec. 7, let the payoff vector be

$$v = \begin{pmatrix} 2 \\ 1 \\ v_3 \\ v_4 \end{pmatrix}.$$

Consider the case of the maximizing player. The various possibilities are indicated in Fig. 2. If $\begin{pmatrix} v_3 \\ v_4 \end{pmatrix}$ is in the interior of region IV then $z = z(\phi)$, $v_3 < z_3$, and $v_4 < z_4$. Hence the player should play on each interior state. If $\begin{pmatrix} v_3 \\ v_4 \end{pmatrix}$ is in the interior region II or its dotted boundary, then the smallest lower semimartingale greater than v is the point on the lower boundary of region I vertically above

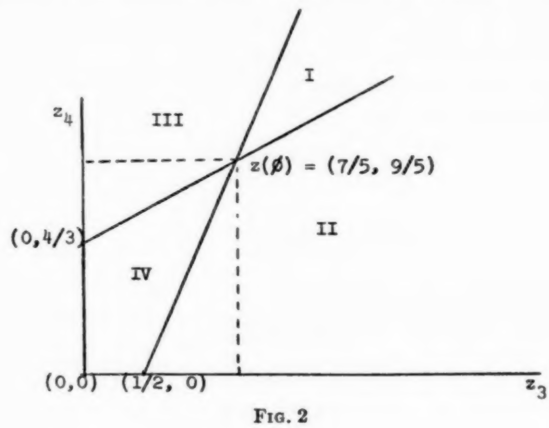


FIG. 2

v. Thus $z_3 = v_3$, $z_4 > v_4$. The player should stop on 3 and play on 4 in this region. Similarly, in region III he should stop on 4 and play on 3. If $\binom{v_3}{v_4}$ is in region I, then $z = v$ and he should not play on any state.

9. Games with a fee for each play. The results of Sec. 8 can be extended to a game in which the player must pay a fee c_i if he wishes to continue playing in interior state i. Alternatively, we may think of c_i as the cost of carrying out an additional experiment. Let c be the column vector which is 0 in B and has components c_i in I. Then by an immediate extension of the previous argument, the vector z giving the values of the various states satisfies

$$z = \max(v, Pz - c). \tag{3}$$

Let d be the column vector such that d_i is the expected cost to reach the boundary from state i. It can be shown that if we take the matrix $\mathscr{I} - P$, where $\mathscr{I}$ is the identity matrix, truncate it to the $s \times s$ matrix obtained by eliminating the boundary states, and take its inverse, then the ijth entry of the resulting matrix gives the expected number of times the process will be in state j if it starts in state i. (See [3], Chapter VII, Sec. 4.) This matrix multiplied into the truncated c-vector gives the truncated d-vector. Remembering that both vectors are 0 in B, we see that

$$(\mathscr{I} - P)\, d = c.$$

Hence

$$Pd = d - c. \tag{4}$$

Since d is a fixed vector, we have from (3) that

$$z + d = \max(v + d, Pz - c + d)$$

and from (4) we see that

$$z + d = \max(v + d, P(z + d)).$$

But this is the problem we solved above. The vector $z + d$ is the least lower semimartingale greater than $v + d$. Thus the value of the game is given by the vector z that is found: First we find the least lower semimartingale greater than $v + d$, then we subtract d.

Thus the game with the cost vector c is strategically equivalent to a costless game in which the payoff vector v has added to it the expected cost of reaching the boundary.

10. Application to discrete subharmonic theory. Consider the lattice of points in the plane of the form (m, n) where m and n are integers. A random walk in the plane is a process which moves from (x, y) to $(x + 1, y)$, $(x - 1, y)$, $(x, y + 1)$, $(x, y - 1)$ with equal probabilities.

Let B and I be finite sets of lattice points such that from any point of I a random walk can reach a point of B but cannot reach any point not in $B \cup I$ without going through B. Then B is called a boundary set and I an interior set.

Consider a boundary set B and interior set I. Assume that boundary values $v(j, k)$ are given on B. Then there is a unique lattice function f defined on $B \cup I$ having the property that

$$f(j,k) = 1/4f(j+1,k) + 1/4f(j,k+1) + 1/4f(j-1,k) + 1/4f(j,k-1), \qquad (j,k) \varepsilon I,$$

and

$$f(j,k) = v(j,k), \qquad (j,k) \varepsilon B.$$

This function provides the discrete analogue for the solution of the Dirichlet problem; the function f is a discrete harmonic function. One should ask the corresponding problem for discrete subharmonic functions. That is, a function f is a discrete subharmonic function with prescribed boundary values if

$$f(j,k) \geqq 1/4f(j+1,k) + 1/4f(j,k+1) + 1/4f(j-1,k) + 1/4f(j,k-1), \qquad (j,k) \varepsilon I,$$

$$f(j,k) = v(j,k), \qquad (j,k) \varepsilon B.$$

In this case the solution would not be unique.

The random walk in $I \cup B$ forms an absorbing Markov chain. Assume that the boundary values are nonnegative. The harmonic function f is given by the vector $z(\phi)$. The subharmonic functions are the semimartingale vectors. Thus the set of all subharmonic functions forms a convex polyhedron and each such function may be represented in terms of a finite number of basic semimartingales. Each basic semimartingale $z(T)$ is simply the unique solution for the Dirichlet problem for boundary $B \cup T$ with the given values on B and 0 on T. Thus the set of all subharmonic functions for a given boundary B may be represented as convex combinations of the harmonic functions for the boundary sets $B \cup T$.

We have reason to believe that these results obtained for discrete subharmonic functions will lead to analogous results for ordinary subharmonic functions.

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, New York, 1953.
[2] J. L. Snell, "Application of martingale system theorems," *Trans. Amer. Math. Soc.*, Vol. 73 (1952), pp. 293–312.
[3] J. G. Kemeny, G. L. Thompson, and J. L. Snell, *Introduction to Finite Mathematics*, New Jersey, 1957.

NOTE ON SUFFICIENT STATISTICS AND TWO-STAGE PROCEDURES

By S. G. Ghurye

University of Chicago

0. Introduction. This note is the result of an attempt to discover problems in which one can apply the two-stage procedure used by Stein [1] for tests regarding the mean of a normal population. One such problem, that of testing for a location parameter of an exponential population, was found to be easily soluble along the lines of Stein's work. An investigation of the problem of optimum statistics for such procedures was also undertaken, and partial solutions, given in Sec. 2, were found. In this connection, the author would like to thank the referee for his useful comments.

1. Testing for a location parameter of a distribution. Throughout this paper $F(x)$ will be a one-dimensional c.d.f. with at least two points of increase. Further, $\{Y_n\}$ will always denote a sequence of independent random variables having a common c.d.f. $F(x)$ and $\{X_n\}$ will denote a family of sequences of independent random variables, all elements of any one sequence having a common c.d.f. $F[(x - \theta)/\sigma]$, $-\infty < \theta < \infty, \sigma > 0$. We shall be dealing with statistics or sequences of real and single-valued functions $t(n; x_1, \cdots, x_n)$ and $s(n; x_1, \cdots, x_n)$ of n real variables, $n = 1, 2, \cdots$, about which one or more of the following assumptions will be made as required:

Assumption I. For any integer $n > 0$, any $a > 0$, any real b and any

$$(x_1, \cdots, x_n) \varepsilon R^n,$$

$$t(n; ax_1 + b, \cdots, ax_n + b) = at(n; x_1, \cdots, x_n) + b. \tag{1}$$

Assumption II. Analogously,

$$s(n; ax_1 + b, \cdots, ax_n + b) = as(n; x_1, \cdots, x_n). \tag{2}$$

Assumption III. There exists a positive, nondecreasing and unbounded sequence $k(n)$ such that

$$\Pr \{t(n; Y_1, \cdots, Y_n) \leqq x/k(n)\} = G(x) \tag{3}$$

is independent of n. Without loss of generality, we may assume $k(1) = 1$.

Assumption IV. The random variables $t(n; Y_1, \cdots, Y_n)$ and $s(n; Y_1, \cdots, Y)_n$ are stochastically independent.

Assumption V. There exists a positive integer m, such that for any $n > m$, $t(n; x_1, \cdots, x_n)$ is a function only of m, n, $t(m; x_1, \cdots, x_m)$ and $x_{m+1}, \cdots, x_n$.

Received January 16, 1956; revised March 18, 1957.

ASSUMPTION VI. Let

$$\Pr\{s(n; Y_1, \cdots, Y_n) \leqq x\} = H(x; n). \tag{4}$$

Then $H(0; n) = 0$ for all n.

Now, let $\mathcal{P}$ be a population whose c.d.f. is known to be $F[(x - \theta)/\sigma]$, but θ, σ are unknown, and suppose that it is desired to obtain a test of $H_0: \theta = \theta_0$ against the alternative $\theta > \theta_0$ with the following properties:

(a) The size of the test is to be a prescribed probability α for all $\sigma > 0$;

(b) The power of the test for $\theta = \theta_0 + \delta$, where $\delta > 0$ is a given number, must be not less than a prescribed probability $\beta > \alpha$ for all $\sigma > 0$;

(c) The power $\rightarrow 1$ as $\theta \rightarrow \infty$.

It is easy to see that, if Assumptions I–VI are satisfied, the following procedure, which is essentially that given by Stein, has the required properties:

Choose an m satisfying Assumption V, and let

$$\chi = \text{infimum of all } y \text{ such that } \int_{-\infty}^{\infty} G(yu)\, dH(u; m) \geqq 1 - \alpha; \tag{5}$$

$$\gamma = \begin{cases} \left[\int_{-\infty}^{\infty} G(\chi u)\, dH(u; m) - 1 + \alpha\right] \Big/ \int_{-\infty}^{\infty} \{G(\chi u) - G(\chi u - 0)\}\, dH(u; m) & \\ \qquad\qquad \text{if the denominator} > 0, & \\ 0 \quad \text{otherwise}; & \end{cases} \tag{6}$$

$$P(y) = 1 - \int_{-\infty}^{\infty} G(yu)\, dH(u; m) + \gamma \int_{-\infty}^{\infty} \{G(yu) - G(yu - 0)\}\, dH(u; m); \tag{7}$$

$$\chi' = \text{supremum of all } y \text{ such that } P(y) \geqq \beta; \tag{8}$$

$$\rho = (\chi - \chi')/\delta > 0. \tag{9}$$

Take m independent observations $X_1, \cdots, X_m$ from $\mathcal{P}$, and calculate

$$s_m = s(m; X_1, \cdots, X_m).$$

Let N be such that

$$k(N - 1) < \rho s_m \leqq k(N), \tag{10}$$

except if $\rho s_m < k(m)$, in which case $N = m$.

If $N > m$, take $N - m$ more independent observations $X_{m+1}, \cdots, X_N$ from $\mathcal{P}$, calculate

$$U = \{t(N; X_1, \cdots, X_N) - \theta_0\} k(N)/s_m, \tag{11}$$

and reject H_0 with probability $\phi(U)$, where

$$\phi(u) = \begin{cases} 0, & u < \chi, \\ \gamma, & u = \chi, \\ 1, & u > \chi. \end{cases} \tag{12}$$

The expected sample-size is

$$E(N) = m \Pr \{\rho s_m \leqq k(m)\} + \sum_{r=1}^{\infty} (m + r) \Pr \{k(m + r - 1) < \rho s_m \leqq k(m + r)\} \tag{13}$$
$$= m + \sum_{r=m}^{\infty} \bar{H}\{k(r)\sigma^{-1}\rho^{-1}; m\},$$

where $\bar{H}(x) = 1 - H(x)$.

For computational convenience, we have the inequalities

$$\nu < E(N) < \nu + \epsilon, \tag{14}$$

where

$$\nu = mH\{k(m)\sigma^{-1}\rho^{-1}; m\} + \int_{\sigma\rho u > k(m)} k^{-1}(\sigma\rho u)\, dH(u; m), \tag{15}$$

$$\epsilon = \bar{H}\{k(m)\sigma^{-1}\rho^{-1}; m\}, \tag{16}$$

and $k^{-1}(u)$ is any monotone function of $u > 0$ such that $k\{k^{-1}(n)\} = n$ for every integer $n > 0$.

It may be noted that, if in Assumption III we drop the restriction $k(1) = 1$, then $ck(n)$, with $c > 0$, serves instead of $k(n)$ (with a different G for each c). However from (10) it is easy to see that N is independent of this c. Thus the restriction $k(1) = 1$ does not cause any loss of generality. In the same way, it can be seen that if we substitute $cs(n; x_1, \cdots, x_n)$ for $s(n; x_1, \cdots, x_n)$, $c > 0$, N is unaffected.

Examples in which t and s satisfying the assumptions can be found are provided by the normal distribution, which was discussed in detail by Stein, and the exponential distribution which we shall take up here.

Of the several possible choices for (t, s) in the normal case, Stein considered two, in both of which s^2 is the usual estimate of σ^2. By using a special linear function for t, he was able to obtain a test whose power is independent of σ instead of merely being bounded below by a function independent of σ as required in property (b) above. However, he noted that this procedure "wastes information," and advocated one using the sample mean as t. In fact, the use of any statistic other than the sample mean is wasteful in the sense that it leads to a higher expected sample-size, as we shall see in Sec. 2.

Next, let $\mathcal{P}$ be a population with c.d.f. $F[(x - \theta)/\sigma]$, where

$$F(x) = \begin{cases} 0 & \text{if } x \leqq 0, \\ 1 - e^{-x} & \text{if } x > 0. \end{cases} \tag{17}$$

Let $x_{[1]}, \cdots, x_{[n]}$ denote the rearrangement of numbers $x_1, \cdots, x_n$ in ascending order of magnitudes, and let

$$t_0(n; x_1, \cdots, x_n) = x_{[1]} = \min (x_1, \cdots, x_n). \tag{18}$$

This is a sufficient estimator of θ, and we shall see in the next section that it is the best for use as "t". If corresponding to independent observations $X_1, \cdots, X_n$ from $\mathcal{P}$, we put $Z_1 = X_{[1]}$, and

$$Z_i = X_{[i]} - X_{[i-1]}, \qquad i = 2, 3, \cdots, n,$$

the joint density of $Z_1, \cdots, Z_n$ is

$$(19) \quad f(z_1, \cdots, z_n) = \begin{cases} n!\, \sigma^{-n} \exp\{-n(z_1 - \theta) - (n-1)z_2 - \cdots - z_n\}/\sigma \\ \qquad\qquad\qquad \text{if } z_1 \geqq \theta,\ z_i \geqq 0,\ i = 2, 3, \cdots n, \\ 0 \quad \text{otherwise.} \end{cases}$$

Any function satisfying Assumption II is a function only of the differences of the arguments, and hence only of $Z_2, \cdots, Z_n$. It is thus independent of Z_1 and can be used as "s" if it is positive with probability 1. This is true more generally, as shown by Lemma 2 of the next section. It will be seen (Example 2) that, asymptotically as $\sigma \to \infty$, the best statistic to use as s is

$$(20) \qquad s_0(n; x_1, \cdots, x_n) = \sum_1^n x_i - nt_0(n; x_1, \cdots, x_n),$$

which together with t_0 is sufficient for σ.

For this pair of statistics, we have from (19)

$$(21) \qquad k(n) = n, \qquad G(x) = F(x),$$

$$(22) \qquad H(u; m) = \begin{cases} \int_0^u x^{m-2} e^{-x}\, dx/(m-2)!, & u > 0 \\ 0, & u \leqq 0; \end{cases}$$

$$(23) \qquad \begin{aligned} \gamma &= 0, \qquad \chi = \alpha^{-1/(m-1)} - 1, \\ \delta\rho &= \{\alpha^{-1/(m-1)} - \beta^{-1/(m-1)}\}, \end{aligned}$$

and ν, ϵ are given by

$$(24) \qquad \begin{aligned} \nu &= m \int_0^{mc} u^{m-2} e^{-u}\, du/(m-2)! + \{(m-1)/c\} \int_{mc}^{\infty} u^{m-1} e^{-u}\, du/(m-1)! \\ \epsilon &= \int_{mc}^{\infty} u^{m-2} e^{-u}\, du/(m-2)!, \end{aligned}$$

where $c = (\sigma\rho)^{-1}$.

The values of ν and ϵ were calculated for $\alpha = 0.05 = 1 - \beta$ and $\alpha = 0.01 = 1 - \beta$ and several values of m and δ/σ. These are given in Table 1.

2. Optimum choice of statistics. We shall now prove three preliminary lemmas which enable us to show that if a suitable sufficient estimator of θ exists, it minimizes the expected sample size among all t satisfying the assumptions.

LEMMA 1. *Let Y be a real-valued, one-dimensional random variable, and $f(y)$ a*

TABLE 1

Expected sample size

(The entries are given in the form $\nu + \epsilon$ and imply that $\nu < E(N) < \nu + \epsilon$)

		δ/σ						
	m	0.05	0.10	0.20	0.40	0.60	0.80	1.00
$\alpha = .05$	2	379.0	189.4	94.8	47.4	31.6	23.8	19.1
		+1.0	+1.0	+1.0	+1.0	+0.9	+0.9	+0.9
	4	101.8	50.9	25.5	12.8	8.7	6.4	5.7
		+1.0	+1.0	+1.0	+0.9	+0.8	+0.7	+0.6
	6	81.0	40.5	20.3	10.4	7.6	6.5	6.2
		+1.0	+1.0	+1.0	+0.8	+0.5	+0.3	+0.1
	8	73.6	36.9	18.5	10.0	8.3	8.1	8.0
		+1.0	+1.0	+1.0	+0.7	+0.2	+0.0	+0.0
	10	70.0	35.0	17.6	10.7	10.0	10.0	10.0
		+1.0	+1.0	+0.9	+0.3	+0.0	+0.0	+0.0
	20	63.9	32.3	20.3	20.0	20.0	20.0	20.0
		+1.0	+1.0	+0.1	+0.0	+0.0	+0.0	+0.0
$\alpha = .01$	2	1980.0	990.0	495.0	247.5	165.5	123.8	99.0
		+1.0	+1.0	+1.0	+1.0	+1.0	+1.0	+1.0
	4	218.3	109.1	54.6	27.3	18.2	13.7	11.0
		+1.0	+1.0	+1.0	+1.0	+1.0	+0.9	+0.9
	6	151.0	75.5	37.8	18.9	12.6	9.8	8.2
		+1.0	+1.0	+1.0	+0.9	+0.9	+0.8	+0.7
	8	130.1	65.0	32.6	16.3	11.3	9.3	8.2
		+1.0	+1.0	+1.0	+0.9	+0.7	+0.5	+0.2
	10	120.6	60.3	30.0	15.3	11.3	10.3	10.0
		+1.0	+1.0	+1.0	+0.8	+0.5	+0.2	+0.0
	20	104.0	52.0	26.8	20.0	20.0	20.0	20.0
		+1.0	+1.0	+0.8	+0.0	+0.0	+0.0	+0.0

measurable, real-valued function with the property that for any real x, θ and any $\sigma > 0$,

$$\Pr\{f(\sigma Y + \theta) - \theta \leqq \sigma x\} = \Pr\{f(Y) \leqq x\}. \tag{25}$$

Then if $f(y)$ is strictly monotone, there exists an interval I, open or closed, such that $y \in I \Rightarrow f(y) = y$, and $\Pr\{Y \in I\} = 1$.

Proof. To start with, we note that since the right hand member of (25) is a nondecreasing function of x, $f(y)$ cannot be a decreasing function; for if it were, we would have

$$\Pr\{f(Y) \leqq x\} = \Pr\{f(Y + \theta) - \theta \leqq x\} \geqq \Pr\{f(Y) - \theta \leqq x\} \quad \text{for all } \theta > 0.$$

This implies that the c.d.f. of Y is constant and hence contradicts the assumption that $F(x)$ has at least two points of increase.

Therefore $f(y)$ is increasing, and

$$\{y:f(y) \leqq u\} = \{y:y \leqq f^{-1}(u)\}.$$

Let $h(y) = f^{-1}(y) - y$; then from (25) we obtain

$$\Pr\left\{Y < x + \frac{h(\sigma x + \theta)}{\sigma}\right\} = \Pr\{Y < x + h(x)\}, \tag{26}$$

$$\Pr\left\{Y = x + \frac{h(\sigma x + \theta)}{\sigma}\right\} = \Pr\{Y = x + h(x)\}. \tag{27}$$

Suppose there exists a y_0 such that $h(y_0) \neq 0$. From (26), we get the relation

$$\Pr\left\{Y < y_0 + \frac{h(y_0)}{\sigma}\right\} = \Pr\{Y < y_0 + h(y_0)\}.$$

By letting $\sigma \to 0$ and again $\sigma \to \infty$, we see that

$$\begin{cases} h(y_0) > 0 & \text{implies} \quad \Pr\{Y \leqq y_0\} = 1, \\ h(y_0) < 0 & \text{implies} \quad \Pr\{Y < y_0\} = 0. \end{cases} \tag{28}$$

Hence, if there exist x_0, y_0 such that $h(x_0) < 0$ and $h(y_0) > 0$, we must have $x_0 < y_0$. Consequently, there exist points x_0, y_0 (which may be respectively $-\infty$ and ∞) such that

$$\begin{cases} h(x_0) \leqq 0, \quad h(y_0) \geqq 0, \quad h(y) = 0 \quad \text{for} \quad x_0 < y < y_0, \\ \qquad\qquad \text{and} \quad \Pr\{x_0 \leqq Y \leqq y_0\} = 1. \end{cases} \tag{29}$$

Finally, if $h(y_0) > 0$ we see from (27), by choosing $\sigma = 1$ and θ such that $x_0 - y_0 < \theta < 0$, that

$$\Pr\{Y = y_0\} = \Pr\{Y = y_0 + h(y_0)\} = 0$$

from (29), and similarly, $h(x_0) < 0$ implies $\Pr\{X = x_0\} = 0$. Hence the result.

Next we want to consider two statistics, one of which is sufficient for θ and both of which have θ as a location parameter. More specifically we prove

LEMMA 2. *Let $P(\cdot\,; \theta)$, $-\infty < \theta < \infty$, be a family of probability measures on a countably additive class of subsets of a set Ω of points ω; let $f(\omega)$, $g(\omega)$ be measurable real valued functions on Ω such that for any Borel sets S, T on the real line,*

$$P\{f^{-1}(S + \theta) \cap g^{-1}(T + \theta); \theta\} \underset{\theta}{\equiv} P\{f^{-1}(S) \cap g^{-1}(T); 0\}. \tag{30}$$

If $f(\omega)$ is a sufficient statistic for the family $P(\cdot\,; \theta)$, the random variables $g(\omega) - f(\omega)$ and $f(\omega)$ are stochastically independent.

PROOF. Writing $Pf^{-1}(S)$ to denote $P\{f^{-1}(S) \cap \Omega; 0\}$, we have from (30)

$$P\{f^{-1}(S) \cap \Omega; \theta\} = Pf^{-1}(S - \theta). \tag{31}$$

By the Radon-Nikodym Theorem and the sufficiency of $f(\omega)$, we know that corresponding to each set T, there exists an integrable function $\lambda(T \mid x)$ on the real line such that

$$P\{f^{-1}(S) \cap g^{-1}(T); \theta\} \underset{\theta}{\equiv} \int_S \lambda(T \mid x) \, dPf^{-1}(x - \theta).$$

This gives us

$$\begin{aligned} P\{f^{-1}(S) \cap g^{-1}(T); 0\} &= P\{f^{-1}(S + \theta) \cap g^{-1}(T + \theta); \theta\} \\ &= \int_S \lambda(T + \theta \mid x + \theta) \, dPf^{-1}(x). \end{aligned}$$

It follows that for every set T, we have

$$\begin{aligned} P\{f^{-1}(S) \cap g^{-1}(T); 0\} &= \int_S \lambda(T - x \mid 0) \, dPf^{-1}(x) \\ &= \int_S \mu(T - x) \, dPf^{-1}(x), \end{aligned}$$

so that

$$\Pr\{g(\omega) \varepsilon T \mid f(\omega) = x\} = \mu(T - x) \text{ for a.e. } x \, [Pf^{-1}],$$

and consequently

$$\Pr\{g(\omega) - f(\omega) \varepsilon T \mid f(\omega) = x\} = \mu(T)$$

is independent of x.

Corollary 2.1. *Let $t_j(n; x_1, \cdots, x_n)$, $j = 0, 1$, be functions satisfying Assumption* I *and let $t_0(n; x_1, \cdots, x_n)$ be a sufficient statistic for the family of distributions $\prod_{j=1}^n F(x_j - \theta)$, $-\infty < \theta < \infty$. Then for any n, if $X_1, \cdots, X_n$ are independent random variables having the common* c.d.f *$F[(x - \theta) / \sigma]$, the random variables $t_0(n; X_1, \cdots, X_n)$ and $t_1(n; X_1, \cdots, X_n) - t_0(n; X_1, \cdots, X_n)$ are independent.*

Corollary 2.2. *If $t_0(n; x_1, \cdots, x_n)$ is as in Corollary* 2.1 *and $s(n; x_1, \cdots, x_n)$ is any function satisfying Assumption* II, *the random variables $t_0(n; X_1, \cdots, X_n)$ and $s(n; X_1, \cdots, X_n)$ are independent.*

Lemma 3[1]. *Let t_0, t_1, $X_1, \cdots, X_n$ be as in Corollary* 2.1 *and suppose that t_0, t_1 also satisfy Assumption* III *with respective sequences $k_0(n)$, $k_1(n)$. Then*

$$k_0(n) \geqq k_1(n), \tag{32}$$

the equality holding if and only if

$$\Pr\{t_0(n; X_1, \cdots, X_n) = t_1(n; X_1, \cdots, X_n)\} = 1. \tag{33}$$

Further, for any a, b such that $a < 0 < b$,

$$\begin{aligned} &\Pr\{a < t_0(n; X_1, \cdots, X_n) - \theta < b\} \\ &\qquad \geqq \Pr\{a < t_1(n; X_1, \cdots, X_n) - \theta < b\}. \end{aligned} \tag{34}$$

[1] It may be of interest to compare this with the results of Pitman [3], pp. 401–402.

PROOF. From Assumptions I and III, it follows that

$$\Pr\{t_j(n; X_1, \cdots, X_n) \leqq x\} = F\left\{\frac{(x - \theta)}{\sigma} k_j(n)\right\}, \qquad j = 0, 1. \tag{35}$$

Let $f(z) = \int e^{izx}\, dF(x)$, $\quad c_n = k_0(n)/k_1(n)$, and

$$g\{z/k_0(n)\} = E\{e^{iz[t_1(n;\, Y_1, \cdots, Y_n) - t_0(n;\, Y_1, \cdots, Y_n)]}\}.$$

Then from Corollary 2.1 and (35), we get

$$f(c_n z) = f(z)g(z). \tag{36}$$

First suppose $f(z_0) = 0$ for some z_0. Then $f(c_n^r z_0) = 0$ for every r, and hence $c_n \nless 1$. On the other hand, if $f(z) \neq 0$ for all z, $g(z) = f(c_n z)/f(z)$ is a characteristic function. Hence $f(c_n^r z)/f(z) = g(z)g(c_n z) \cdots g(c_n^{r-1} z)$, $r = 1, 2, \cdots$, is a sequence of characteristic functions. If $c_n < 1$, $\lim_{r\to\infty} f(c_n^r z)/f(z) = 1/f(z)$ for every finite z, and since the limit is continuous in z, it is a characteristic function. But $f(z)$ is also a characteristic function. This implies $|f(z)| = 1$ for all z, which is impossible if F has more than one point of increase. Consequently, $c_n \geqq 1$, that is to say (32) holds.

It follows from (36) that $c_n = 1$ if and only if $g(z) = 1$, in which case (33) follows on account of Assumption I.

Finally, (34) is an immediate consequence of (32) and (35).

THEOREM 1. *Let t_1 and s be statistics satisfying Assumptions* I–VI, *and let t_0 be a statistic, satisfying Assumptions* I, III *and* V, *which is sufficient for the family of distributions $\Pi_1^n F(x_j - \theta)$, $-\infty < \theta < \infty$. Let N_i denote the sample-size in the two-stage procedure using (t_i, s), $i = 0, 1$, with the same m, α, β, and δ. Then*

$$E(N_0) \leqq E(N_1) \quad \textit{for all} \quad \sigma, \tag{37}$$

the equality holding for all σ if and only if

$$\Pr\{t_0(n; X_1, \cdots, X_n) = t_1(n; X_1, \cdots, X_n)\} = 1 \textit{ for all } n \geqq m.$$

PROOF. The hypotheses of the theorem and Corollary 2.2 enable us to use (t_0, s) for the procedure described in the previous section. With the notation used there we have from (30), $G_i(x) = F(x)$, $i = 0, 1$, and from (9),

$$\rho_0 = \rho_1. \tag{38}$$

From (13), we get

$$E(N_i) = m + \sum_{r=m}^{\infty} \bar{H}\{k_i(r)\sigma^{-1}\rho_i^{-1}; m\},$$

and using (32) and (38) the result follows.

REMARK. Theorem 1 solves part of the problem of optimization of the two-sample procedure by showing that if a suitable sufficient estimator of θ exists, it is the best "t" to use. This leaves us with the problem of choosing "s". We shall see that in the case of the normal and exponential distributions the best pair (t, s) to use, asymptotically as $\sigma \to \infty$, is the pair of sufficient statistics (t_0, s_0).

LEMMA 4[2] *Let $s_0(n; x_1, \cdots, x_n)$ and $s(n; x_1, \cdots, x_n)$ be statistics satisfying Assumptions* II *and* VI, *and let t_0 be a statistic satisfying Assumption* I. *Let (t_0, s_0) be sufficient for the family $\Pi_1^n F[(x_i - \theta)/\sigma]$, $-\infty < \theta < \infty$, $\sigma > 0$. Then $t_0(n; X_1, \cdots, X_n)$, $s_0(n; X_1, \cdots, X_n)$, and $s(n; X_1, \cdots, X_n)/s_0(n; X_1, \cdots, X_n)$ are mutually independent.*

PROOF. This result can be proved formally along the lines used for Lemma 2, but this seems hardly necessary, and only an outline in terms of conditional probabilities will be given.

Let $u(n; x_1, \cdots, x_n) = s(n; x_1, \cdots, x_n)/s_0(n; x_1, \cdots, x_n)$, and note that u is invariant under the transformation $x_i \rightarrow \sigma x_i + \theta$, $i = 1, \cdots, n$.

For almost all a and b,

$$\Pr\{u(n; X_1, \cdots, X_n) \,\varepsilon\, S \mid t_0(n; X_1, \cdots, X_n) = a, \quad s_0(n; X_1, \cdots, X_n) = b\} \tag{39}$$

is independent of (θ, σ) and equals

$$\Pr\{u(n; Y_1, \cdots, Y_n) \,\varepsilon\, S \mid t_0(n; Y_1, \cdots, Y_n) = a, \quad s_0(n; Y_1, \cdots, Y_n) = b\}, \tag{40}$$

using notation indicated at the beginning of Sec. 1. On the other hand, from the hypotheses of the lemma, (39) also equals

$$\Pr\left\{u(n; Y_1, \cdots, Y_n) \,\varepsilon\, S \mid t_0(n; Y_1, \cdots, Y_n) = \frac{a - \theta}{\sigma}, \quad s_0(n; Y_1, \cdots, Y_n) = \frac{b}{\sigma}\right\}. \tag{41}$$

From the equality of (40) and (41), it follows that the conditional distribution of u is independent of the conditioning values of t_0 and s_0, so that u is stochastically independent of (t_0, s_0). But t_0 and s_0 are mutually independent, since Corollary 2.2 applies. Hence the result.

This lemma can be used to compare the relative merits of s_0 and any other s asymptotically as $\sigma \rightarrow \infty$. Let us assume that the hypotheses of Lemma 4 are satisfied, that F is continuous and that t_0 also satisfies Assumptions III and V. Then we know that t_0 is the best statistic to use as "t", and both s_0 and s are eligible as the "s" statistic. Let

$$J(u) = \Pr\{s(m; X_1, \cdots, X_m) \leqq u s_0(m; X_1, \cdots, X_m)\}, \tag{42}$$

and $H(u)$, $H_0(u)$ denote the c.d.f.'s of s and s_0 respectively. It will be understood that we have the same m throughout the discussion. We already know

$$\Pr\{t_0(n; X_1, \cdots, X_n) \leqq \theta + \sigma x\} = F\{x k_0(n)\}. \tag{43}$$

---

[2] My attention has been drawn to the fact that a general result of the type of those given in Lemmas 2 and 4 has been previously given, for boundedly complete sufficient statistics, by D. Basu [*Sankhya*, vol. 15 (1955), pp. 377–380.]

Let

$$M(y) = \int_0^\infty F(yu)\, dH_0(u). \tag{44}$$

Then

$$H(v) = \int_0^\infty H_0(v/u)\, dJ(u) \tag{45}$$

by Lemma 4, and we get

$$\int_0^\infty F(yu)\, dH(u) = \int_0^\infty M(yu)\, dJ(u). \tag{46}$$

From (5), (8), and (9) on account of continuity, we have

$$\rho = (\chi - \chi')/\delta, \qquad \rho_0 = (\chi_0 - \chi_0')/\delta, \tag{47}$$

and

$$\begin{cases} M(\chi_0) = 1 - \alpha = \displaystyle\int_0^\infty M(\chi u)\, dJ(u), \\ M(\chi_0') = 1 - \beta = \displaystyle\int_0^\infty M(\chi' u)\, dJ(u). \end{cases} \tag{48}$$

Hence,

$$(\chi_0 - \chi_0') = M^{-1}\left\{\int_0^\infty M(\chi u)\, dJ(u)\right\} - M^{-1}\left\{\int_0^\infty M(\chi' u)\, dJ(u)\right\}. \tag{49}$$

Now, from (14) and (15) we know that $E(N)$, $E(N_0) \to \infty$ as $\sigma \to \infty$, and

$$\frac{E(N_0)}{E(N)} \cong \frac{\displaystyle\int_0^\infty k_0^{-1}(\sigma\rho_0 u)\, dH_0(u)}{\displaystyle\int_0^\infty k^{-1}(\sigma\rho u)\, dH(u)}.$$

Suppose

$$k(u) = u^{1/c}, \tag{50}$$

where c is a constant $\geqq 1$. (This is the case in the normal and exponential populations.) Then

$$\frac{E(N_0)}{E(N)} \cong \frac{\rho_0^c \displaystyle\int_0^\infty u^c\, dH_0(u)}{\rho^c \displaystyle\int_0^\infty u^c\, dH(u)} = \frac{\rho_0^c}{\rho^c \displaystyle\int_0^\infty u^c\, dJ(u)}, \tag{51}$$

by (45).

Therefore

$$\rho\left\{\int_0^\infty u^c\, dJ(u)\right\}^{1/c} > \rho_0$$

implies that, asymptotically as $\sigma \to \infty$, $E(N_0) < E(N)$. However, since

$$\left\{\int_0^\infty u^c\, dJ(u)\right\}^{1/c} > \int_0^\infty u\, dJ(u),$$

if we can show that

$$(52) \qquad (\chi - \chi')\int_0^\infty u\, dJ(u) > \chi_0 - \chi_0',$$

this implies that asymptotically (t_0, s_0) is the best, or minimum-expected-sample-size, pair among those satisfying the initial assumptions.

We shall now prove (52) to hold in the two cases that matter. As previously noted, s_0 and as_0, where a is a constant >0, are equivalent statistics for our purpose, and hence in what follows we shall only consider as alternative candidates, statistics s which are not constant multiples of s_0; in other words, we assume that $J(u)$ has at least two points of increase.

EXAMPLE 1. Let $F(x) = \int_{-\infty}^{x} e^{-(u^2/2)}\, du/\sqrt{2\pi}$, and assume $\alpha < 0.5 < \beta$. $M(y)$ is Student's distribution, so that $M(0) = 0.5$. Hence

$$(53) \qquad \chi_0' < 0 < \chi_0 \quad \text{and} \quad \chi' < 0 < \chi.$$

Further, $M(y)$ is concave or convex according as $y > 0$ or <0, and therefore $M^{-1}\int_0^\infty M(yu)\, dJ(u) \lesseqgtr y\int_0^\infty u\, dJ(u)$ according as $y \gtrless 0$. From (53) and (49), (52) follows.

EXAMPLE 2. Let $F(x)$ be given by (17). Then

$$M(y) = \begin{cases} 0 & \text{if } y \leqq 0 \\ 1 - (1+y)^{-\mu} & \text{if } y > 0, \end{cases} \qquad \text{where } \mu = m - 1 \geqq 1.$$

Consequently, all χ's are positive. Now let

$$(54) \qquad \begin{aligned} f(y) &= y\int_0^\infty u\, dJ(u) - M^{-1}\int_0^\infty M(yu)\, dJ(u) \\ &= y\int_0^\infty u\, dJ(u) - \left\{\int_0^\infty (1+yu)^{-\mu}\, dJ(u)\right\}^{-1/\mu} - 1. \end{aligned}$$

Then

$$f'(y) = \int_0^\infty u\, dJ(u) - \int_0^\infty u(1+yu)^{-\mu-1}\, dJ(u) \cdot \left\{\int_0^\infty (1+yu)^{-\mu}\, dJ(u)\right\}^{-(\mu+1)/\mu}.$$

It is easily seen that $f'(y) > 0$ for $y > 0$; because the sign of $f'(y)$ is the same as that of

$$\left\{\int_0^\infty (1 + yu)^{-\mu}\, dJ(u)\right\}^{(\mu+1)/\mu} \int_0^\infty u\, dJ(u) - \int_0^\infty u(1 + yu)^{-\mu-1}\, dJ(u)$$

which is

$$> \left\{\int (1 + yu)^{-\mu}\, dJ(u)\right\}\left\{\int (1 + yu)^{-1}\, dJ(u)\right\} \int u dJ(u)$$

$$- \int u(1 + yu)^{-\mu-1}\, dJ(u)$$

$$> \int (1 + yu)^{-\mu}\, dJ(u) \int u(1 + yu)^{-1}\, dJ(u) - \int u(1 + yu)^{-\mu-1}\, dJ(u)$$

$$> 0,$$

since u and $(1 + yu)^{-1}$ are monotone in opposite directions for $y > 0$, and the same is true of $u(1 + yu)^{-1}$ and $(1 + yu)^{-\mu}$. Consequently, $f(y)$ is an increasing function of $y > 0$; (52) follows from (49), and asymptotically as $\sigma \to \infty$, (t_0, s_0) is the best pair of statistics to use.

REFERENCES

[1] C. Stein, "A two-sample test of a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

[2] S. G. Ghurye and Herbert Robbins, "Two-stage procedures for estimating the difference between means," *Biometrika*, Vol. 41 (1954), pp. 146–152.

[3] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.

SAMPLING VARIANCES OF ESTIMATES OF COMPONENTS OF VARIANCE

By S. R. Searle

New Zealand Dairy Board, Wellington, N. Z.[1]

1. Outline. In earlier work (4) matrix methods have been developed for obtaining the sampling variances of estimates of components of variance. These rely on the fact that if $y = \mathbf{x}'F\mathbf{x}$ is a function of variables $\mathbf{x}$, having a multinormal distribution with variance-covariance matrix V, then the variance of y is given by

$$\text{var}(y) = 2 \text{ tr } (VF)^2. \tag{1}$$

The use of the method was demonstrated by obtaining for the case of a 1-way classification with unequal numbers in the sub-classes, the sampling variances of the estimates of variance components, as summarized in (1); it was then extended to the sampling variances of estimates of components of covariance.

The present paper makes further use of this matrix technique to obtain the sampling variances of estimates of components of variance from data in a 2-way classification having unequal sub-class numbers. The model assumed is Eisenhart's Model II, [2], and the method of estimating the components is taken to be Henderson's Method 1, [3].

2. Model and analysis of variance. The observations x_{ijk} are taken as having the linear model

$$x_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk},$$

with $k = 1 \cdots n_{ij}$, $i = 1 \cdots a$, and $j = 1 \cdots b$. μ is a general mean, A_i and B_j are main effects, $(AB)_{ij}$ is an interaction and ϵ_{ijk} is residual error. Under the assumptions of the model, all terms (except μ) are taken as being normally distributed, with zero means, and variances σ_a^2, σ_b^2, σ_{ab}^2, and σ_ϵ^2, which we will write as α, β, γ and ϵ respectively.

For a sample of N observations in N' cells of this 2-way classification an analysis of variance can be written as

Term	d-f	Sums of Square
Between A classes	$a - 1$	$T_a - T_f = S_a$
Between B classes	$b - 1$	$T_b - T_f = S_b$
Interaction $A \times B$	$N' - a - b + 1$	$T_{ab} - T_a - T_b + T_f = S_{ab}$
Residual	$N - N'$	$T_0 - T_{ab} = S_w$
Total	$N - 1$	$T_0 - T_f$

where the T's are uncorrected sums of squares. With $n_{i.} = \sum_j n_{ij}$, and $n_{.j}$

Received December 3, 1956; revised September 5, 1957.

[1] The author's present address is Animal Husbandry Dept., Cornell University, Ithaca, New York.

$= \sum_i n_{ij}$, and using customary notation for means,

$$T_a = \sum_i n_{i.} \bar{x}_{i..}^2, \qquad T_b = \sum_j n_{.j} \bar{x}_{.j.}^2,$$

$$T_{ab} = \sum_i \sum_j n_{ij} \bar{x}_{ij.}^2, \quad T_f = N\bar{x}_{...}^2, \qquad \text{and } T_0 = \sum_i \sum_j \sum_k x_{ijk}^2.$$

We may note in passing that not all the expressions in the "sums of squares" column are in fact sums of squares, notably the interaction term. It would be more correct to label this column "quadratic forms" but the terminology "sums of squares" has historical precedence and will be retained.

Henderson's first method [3] for estimating the components of variance is to equate each of the first four lines in the above analysis to its expected value. Denoting the resulting estimates of α, β, γ, and ϵ as $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\epsilon}$, the equations for obtaining them are

$$(2)\quad \left\{\begin{aligned} T_a - T_f &= S_a = (N - k_1)\hat{\alpha} + (k_{12} - k_2)\hat{\beta} \\ &\qquad + (k_{12} - k_3)\hat{\gamma} + (a - 1)\hat{\epsilon} \\ T_b - T_f &= S_b = (k_{21} - k_1)\hat{\alpha} + (N - k_2)\hat{\beta} \\ &\qquad + (k_{21} - k_3)\hat{\gamma} + (b - 1)\hat{\epsilon} \\ T_{ab} - T_a - T_b + T_f &= S_{ab} = (k_1 - k_{21})\hat{\alpha} + (k_2 - k_{12})\hat{\beta} \\ &\qquad + (N - k_{12} - k_{21} + k_3)\hat{\gamma} \\ &\qquad + (N' - a - b + 1)\hat{\epsilon} \end{aligned}\right\}$$

$$(3)\quad T_0 - T_{ab} = S_w = (N - N')\hat{\epsilon}$$

where the k's are functions of the n_{ij}'s, namely

$$k_{12} = \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} \qquad k_{21} = \sum_j \frac{\sum_i n_{ij}^2}{n_{.j}},$$

$$k_1 = \frac{1}{N}\sum_i n_{i.}^2, \qquad k_2 = \frac{1}{N}\sum n_{.j}^2, \qquad \text{and } k_3 = \frac{1}{N}\sum n_{ij}^2.$$

3. Variances required. In the analysis of variance S_w has a χ^2-distribution with $N - N'$ degrees of freedom. Hence, from Eq. (3) the variance of $\hat{\epsilon}$ is

$$\sigma_{\hat{\epsilon}}^2 = \frac{2\epsilon^2}{N - N'}.$$

Using (3), Eqs. (2) give $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ as linear functions of S_a, S_b, S_{ab}, and $\hat{\epsilon}$. But S_w, and hence $\hat{\epsilon}$, is distributed independently of S_a, S_b, and S_{ab}. Hence the variances and covariances of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ can be obtained as linear functions of $\sigma_{\hat{\epsilon}}^2$ and the variances and covariances of S_a, S_b, and S_{ab}. By the nature of the S's it is easier to consider the variances and covariances of T_a, T_b, T_{ab}, and T_f.

Writing P for the matrix of coefficients of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ in Eqs. (2) these equations can be written as

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} T_a \\ T_b \\ T_{ab} \\ T_f \end{pmatrix} = P \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} + \hat{\epsilon} \begin{pmatrix} a-1 \\ b-1 \\ N'-a-b+1 \end{pmatrix},$$

which we may write as

$$H\mathbf{t} = P\mathbf{v} + \hat{\epsilon}\mathbf{m}.$$

Since $\hat{\epsilon}$ is independent of the terms in $H\mathbf{t}$, the variance-covariance matrix of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, var $(\mathbf{vv}')$, can be expressed in terms of the variance-covariance matrix of the T's, var $(\mathbf{tt}')$, as

(4) $$\text{var}\,(\mathbf{vv}') = P^{-1}[H\,\text{var}\,(\mathbf{tt}')H' + \mathbf{mm}'\sigma_{\hat{\epsilon}}^2]P^{-1'}$$

and

$$\text{cov}\,(\hat{\epsilon}\mathbf{v}) = -P^{-1}\mathbf{m}\sigma_{\hat{\epsilon}}^2 .$$

The unknown term in these expressions is var $(\mathbf{tt}')$ the variance-covariance matrix of T_a, T_b, T_{ab}, and T_f, which we now proceed to obtain, term by term.

4. Matrix definitions and expressions. Let U be a matrix having a one for every element, its order being denoted by subscripts, thus:—

U-matrix (*all elements 1*)	Order
$U_{ij,kl}$	$n_{ij} \times n_{kl}$
$U_{ij,k.}$	$n_{ij} \times n_{k.}$
U_{ij}	$n_{ij} \times n_{ij}$
$U_{i.}$	$n_{i.} \times n_{i.}$
U_N	$N \times N$

Define W-matrices in terms of the U's:

$$W_{ij} = \frac{1}{n_{ij}} U_{ij},$$

$$W_{i.} = \frac{1}{n_{i.}} U_{i.},$$

$$W_{.j} = \frac{1}{n_{.j}} U_{.j}, \text{ and } W_N = \frac{1}{N} U_N .$$

Then C-matrices are defined, of order $N \times N$, whose only non-zero submatrices are W's along the diagonal:

C_a	has	$W_{i.}(i = 1 \cdots a)$	in the diagonal,
C_b	has	$W_{.j}(j = 1 \cdots b)$	in the diagonal,
C_{ab}	has	$W_{ij}(i = 1 \cdots a, j = 1 \cdots b)$,	in the diagonal.

Finally we define D-matrices, the same as C-matrices only having U-matrices instead of W-matrices in their diagonals.

Let $\mathbf{x}'$ be the row vector of the N x_{ijk}'s, arrayed in order, $k = 1 \cdots n_{ij}$, within j-classes, within each i-class; i.e.,

$$\mathbf{x}' = (x_{111} \cdots x_{11n_{11}} \quad x_{121} \cdots x_{12n_{12}} \cdots x_{ab1} \cdots x_{abn_{ab}}).$$

Then if $\mathbf{w}'$ is the vector of the x's arrayed in k-order within i-classes within each j-class, $\mathbf{w}'$ will be a transform of $\mathbf{x}'$, $\mathbf{w}' = \mathbf{x}'R'$, say, where R is an orthogonal elementary operational matrix of order N, of identity matrices I.

The T's can now be expressed in terms of these vectors and matrices:

$$\begin{aligned} T_a &= \mathbf{x}'C_a\mathbf{x}, \\ T_b &= \mathbf{w}'C_b\mathbf{w} = \mathbf{x}'R'C_bR\mathbf{x} = \mathbf{x}'B\mathbf{x}\text{, say.} \\ T_{ab} &= \mathbf{x}'C_{ab}\mathbf{x} = \mathbf{w}'C_{ba}\mathbf{w}, \\ T_f &= \mathbf{x}'U_N\mathbf{x}. \end{aligned}$$

In C_{ab} the W_{ij} in the diagonal are in j-order within i-order; in C_{ba} they are in i-order within j-order.

V, the variance-covariance matrix of the x_{ijk}'s appropriate to $\mathbf{x}'$ can be written as

$$V = J + K,$$

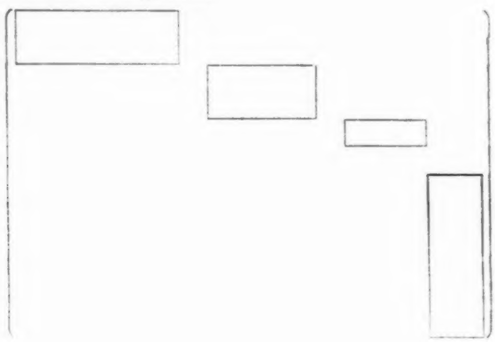
where

$$J = \alpha D_a + (\beta + \gamma)D_{ab} + \epsilon I,$$

and

$$K = \begin{pmatrix} 0 & K_{12} & K_{13} \cdots & K_{1a} \\ K_{21} & 0 & & \\ \vdots & & & \\ K_{a1} & & & 0 \end{pmatrix}$$

with $K_{ii'}$, $i \neq i'$, $i, i' = 1 \cdots a$, of order $n_{i.} \times n_{i'.}$ has all elements zero except those in b rectangular matrices $\beta U_{ij,i'j}$, $j = 1 \cdots b$. These b matrices lie "corner to corner" across $K_{ii'}$ thus:



For example the V matrix for a sample of 7 observations with $n_{11} = 1$, $n_{12} = 2$, $n_{21} = 3$ and $n_{22} = 1$ would be

$$\left(\begin{array}{ccc|cccc} \alpha+\beta+\gamma+\epsilon & \alpha & \alpha & \beta & \beta & \beta & \cdot \\ \alpha & \alpha+\beta+\gamma+\epsilon & \alpha+\beta+\gamma & \cdot & \cdot & \cdot & \beta \\ \alpha & \alpha+\beta+\gamma & \alpha+\beta+\gamma+\epsilon & \cdot & \cdot & \cdot & \beta \\ \hline \beta & \cdot & \cdot & \alpha+\beta+\gamma+\epsilon & \alpha+\beta+\gamma & \alpha+\beta+\gamma & \alpha \\ \beta & \cdot & \cdot & \alpha+\beta+\gamma & \alpha+\beta+\gamma+\epsilon & \alpha+\beta+\gamma & \alpha \\ \beta & \cdot & \cdot & \alpha+\beta+\gamma & \alpha+\beta+\gamma & \alpha+\beta+\gamma+\epsilon & \alpha \\ \cdot & \beta & \beta & \alpha & \alpha & \alpha & \alpha+\beta+\gamma+\epsilon \end{array}\right)$$

This is the variance-covariance matrix appropriate to $\mathbf{x}$: that for $\mathbf{w}$ will be RVR'.

5. Variances and covariances. T_a, T_b, T_{ab}, and T_f have now been expressed in the form $\mathbf{x}'F\mathbf{x}$, and the variance-covariance matrix has also been obtained. The sampling variances of the T's will be found from (1), by evaluating 2 trace $(VF)^2$ for each of them.

5.1. $$\text{var}\,(T_a) = 2\,\text{tr}\,(VC_a)^2.$$

VC_a can be expressed as

$$(VC_a) = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1a} \\ P_{21} & \ddots & & \\ \vdots & & \ddots & \\ P_{a1} & & & P_{aa} \end{pmatrix},$$

where P_{ii} is a column of matrices $x_{ij}U_{ij.i.}$, $(j = 1 \cdots b)$ with

$$x_{ij} = (1/n_{i.})(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon).$$

Similarly $P_{ii'}$ is a column of matrices $w_{ij}U_{ij.i'.}$, $(j = 1 \cdots b)$, with

$$w_{i'j} = (n_{i'j}/n_{i'.})\beta.$$

VC_a has here been partitioned into P-matrices, which themselves have been partitioned into sub-matrices of the U-type. Trace $(VC_a)^2$ will therefore depend on two properties of these U matrices, that

$$U_{ij.pq}U_{pq.rs} = n_{pq}U_{ij.rs} \tag{5}$$

and

$$\text{tr}\,(U_{ij}) = n_{ij}\,.$$

Hence

$$\text{tr}\,(U_{ij.pq}U_{pq.ij}) = n_{ij}n_{pq}\,. \tag{6}$$

Using these results we have

$$\begin{aligned}\operatorname{tr}(VC_a)^2 &= \sum_i \sum_{i'} \operatorname{tr}(P_{ii'}P_{i'i}) \\ &= \sum_i \operatorname{tr}(P_{ii}^2) + \sum_i \sum_{i' \neq i} \operatorname{tr}(P_{ii'}P_{i'i}) \\ &= \sum_i \sum_j n_{ij}x_{ij} \sum_j n_{ij}x_{ij} + \sum_i \sum_{i' \neq i} \sum_j n_{ij}w_{i'j} \sum_j n_{i'j}w_{ij}.\end{aligned}$$

On substituting for x_{ij} and w_{ij} this gives

$$\tfrac{1}{2}\operatorname{var}(T_a) = \sum_i \Big[\sum_j n_{ij}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}\Big]^2 + \sum_i \sum_{i' \neq i} \frac{\big(\sum_j n_{ij}n_{i'j}\big)^2}{n_{i.}n_{i'.}}\beta^2.$$

5.2. $$\operatorname{var}(T_{ab}) = 2\operatorname{tr}(VC_{ab})^2.$$

V and C_{ab} are such that their product can be written as

$$VC_{ab} = L + K,$$

where K is as in V, and

$$L = \alpha D_a + (\beta + \gamma)D_{ab} + \epsilon C_{ab}.$$

Hence,

$$VC_{ab} = V + \epsilon(C_{ab} - I). \tag{7}$$

Since V and C_{ab} are symmetric, VC_{ab} is also, and hence squaring (7) gives

$$(VC_{ab})^2 = V^2 + \epsilon^2(C_{ab}^2 - I).$$

Hence,

$$\tfrac{1}{2}\operatorname{var}(T_{ab}) = \operatorname{tr} V^2 + \epsilon^2(\operatorname{tr} C_{ab}^2 - \operatorname{tr} I)$$

$$= \sum_i \sum_j n_{ij}[(\alpha + \beta + \gamma + \epsilon)^2 + (n_{ij} - 1)(\alpha + \beta + \gamma)^2 + (n_{i.} - n_{ij})\alpha^2 + (n_{.j} - n_{ij})\beta^2] + \epsilon^2\Big[\sum_i \sum_j n_{ij}n_{ij}\frac{1}{n_{ij}^2} - N\Big],$$

which reduces to

$$\tfrac{1}{2}\operatorname{var}(\mathrm{T}_{ab}) = \sum_i \sum_j n_{ij}\,[n_{ij}(\alpha + \beta + \gamma + \epsilon/n_{ij})^2 + (n_{i.} - n_{ij})\alpha^2 + (n_{.j} - n_{ij})\beta^2].$$

5.3. $$\operatorname{var}(T_f) = 2\operatorname{tr}(VW_N)^2.$$

Similar to the form of the P-matrices in 5.2, VW_N can be expressed as a column of matrices $y_{ij}U_{ij.N}$, $(i = 1 \cdots a, j = 1 \cdots b)$, where

$$y_{ij} = (n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/N.$$

Hence,

$$\text{tr}\,(VW_N)^2 = \sum_i \sum_j n_{ij} y_{ij} \sum_i \sum_j n_{ij} y_{ij},$$

giving

$$\tfrac{1}{2}\,\text{var}\,(T_f) = [\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon(/N]^2.$$

5.4. In general, for any two square matrices of the same order, A and B say, it can be shown that $\text{tr}\,(A + B)^2 = \text{tr}\,A^2 + \text{tr}\,B^2 + 2\,\text{tr}\,AB$. If then, $\hat{a}$ and $\hat{b}$ are two function of the same set of variables such that $\text{var}\,(\hat{a}) = 2\,\text{tr}\,A^2$, and $\text{var}\,(\hat{b}) = 2\,\text{tr}\,B^2$, it follows at once that

$$\text{(8)} \qquad \text{cov}\,(\hat{a}\hat{b}) = 2\,\text{tr}\,AB = 2\,\text{tr}\,BA.$$

This result will be used for obtaining the covariances among T_a, T_b, T_{ab}, and T_f.

5.5.
$$\text{cov}\,(T_a, T_{ab}) = 2\,\text{tr}\,(VC_a)(VC_{ab}).$$

In 5.1 VC_a has been partitioned into P_{ii}'s and $P_{ii'}$'s. If VC_{ab}, expressed as $L + K$ in 5.2 is partitioned in the same manner, into L_{ii}'s and $K_{ii'}$'s, then

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_{ab}) = \sum_i \sum_{l=1}^{n_{i.}} (\text{inner product of } l\text{'th row of } P_{ii} \text{ and } l\text{ th column of } L_{ii})$$
$$+ \sum_i \sum_{i' \neq i} \sum_{l=1} (\text{inner product of } l\text{'th row of } P_{ii'} \text{ and } l\text{ th column of } K_{i'i})$$

and after substitution this reduces to

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_{ab}) = \sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)^2/n_{i.}$$
$$+ \beta^2 \sum_i \sum_j n_{ij}^2(n_{.j} - n_{ij})/n_{i.}\,.$$

5.6.
$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_f) = \text{tr}\,(VW_N)(VC_a).$$

Using 5.1 and 5.4, and Eq. (6), this can be expressed as

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_f) = \sum_i \sum_j n_{ij} y_{ij} \,(\sum_j n_{ij} x_{ij} + \sum_j \sum_{i' \neq i} n_{i'j} w_{ij}).$$

which on substitution for the x's, y's and w's, reduces to

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_f) = \sum_i \sum_j \frac{n_{ij}}{N}\,(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)$$
$$\cdot \left(n_{i.}\alpha + \beta \frac{\sum_j n_{.j} n_{ij}}{n_{i.}} + \gamma \frac{\sum_j n_{ij}^2}{n_{i.}} + \epsilon\right).$$

5.7.
$$\tfrac{1}{2}\,\text{cov}\,(T_{ab}, T_f) = \text{tr}\,(VW_N)(VC_{ab})$$
$$= \sum_i \sum_j n_{ij} y_{ij}\,[\sum \text{terms in } ij\text{'th column of } V + \epsilon(C_{ab} - I)]$$
$$= \sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2/N.$$

5.8. In Eq. (8) it is required that $\hat{a}$ and $\hat{b}$ be functions of the same set of variables; therefore, in terms of paragraph 4, the covariance of T_a and T_b must be expressed as

$$\operatorname{cov}(T_a, T_b) = 2 \operatorname{tr} (VC_a)(VB).$$

This covariance is a little more cumbersome to evaluate than previous ones; the method used is essentially a generalization of earlier paragraphs.

B is the same form as V, (4.5) but with matrices $(1/n_{.j})U_{ij}$ in the diagonal $j = 1 \cdots b$, for $i = 1 \cdots a$, and with $K_{ii'}$-matrices having terms

$$(1/n.j)U_{ij,i'j}.$$

Now partition V into matrices $(V)_{ij:kl}$ of order $n_{ij} \times n_{kl}$, there being four different forms of this matrix according as k and l are equal or not equal to i and j respectively, namely:

$$\begin{aligned}
(V)_{ij:ij} &= (\alpha + \beta + \gamma)U_{ij} + \epsilon I; & & \\
(V)_{ij:il} &= \alpha U_{ij,il} & &\text{for } l \neq j; \\
(V)_{ij:kj} &= \beta U_{ij,kj} & &\text{for } k \neq l; \\
(V)_{ij:kl} &= 0.U_{ij,kl} \text{ a zero matrix,} & &\text{for } k \neq i,\, l \neq j.
\end{aligned}$$

B can be partitioned similarly for $k \neq i$ and $l \neq j$:

$$\begin{aligned}
(B)_{ij:ij} &= \frac{1}{n_{.j}} U_{ij}, \\
(B)_{il:ij} &= 0.U_{il,ij}, \\
(B)_{kj:ij} &= \frac{1}{n_{:j}} U_{kj,ij}, \\
(B)_{kl:ij} &= 0.U_{kl,ij}.
\end{aligned}$$

Consider now the identity

$$(9) \qquad (VB)_{pq:tu} = \sum_f \sum_g V_{pq:fg} B_{fg:tu},$$

whose right-hand side can be expanded as

$$(V)_{pq:tu}(B)_{tu:tu} + \sum_{f \neq t} (V)_{pq:fu}(B)_{fu:tu} + \sum_{g \neq u} (V)_{pq:tg}(B)_{tg:tu} + \sum_{f \neq t} \sum_{g \neq u} (V)_{pq:fg}(B)_{fg:tu}$$

or as

$$(V)_{pq:pq}(B)_{pq:tu} + \sum_{f \neq p} (V)_{pq:fq}(B)_{fq:tu} + \sum_{g \neq q} (V)_{pq:pg}(B)_{pg:tu} + \sum_{f \neq p} \sum_{g \neq q} (V)_{pq:fg}(B)_{fg:tu}.$$

These expressions are true for any values of the subscripts.

Applying this identity to the partitioned forms of V and B given above, and using the principle of (5) in 5.1 gives

$$(VB)_{ij:ij} = n_{ij}(\alpha + \beta + \gamma)/n_{.j}U_{ij} + \epsilon/n_{.j}U_{ij} + \beta \sum_{k \neq i} \frac{1}{n_{.j}} U_{ij,kj}U_{kj,ij}$$
$$= (n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/n_{.j}U_{ij} = b_{ij}U_{ij}, \quad \text{say.}$$

Similarly, for $r \neq i$, and $s \neq j$,

$$(VB)_{is:ij} = \alpha \frac{n_{ij}}{n_{.j}} U_{is,ij} = b'_{ij}U_{is,ij}, \quad \text{say,}$$

$$(VB)_{rj:ij} = (n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/n_{.j}U_{rj,ij} = b_{ij}U_{rj,ij},$$

$$(VB)_{rs:ij} = \alpha \frac{n_{rj}}{n_{.j}} U_{rs,ij} = b'_{rj}U_{rs,ij}.$$

Likewise:

$$(VC_a)_{ij:ij} = (n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}U_{ij} = a_{ij}U_{ij}, \quad \text{say,}$$

$$(VC_a)_{ij:is} = (n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}U_{ij,is} = a_{ij}U_{ij,is},$$

$$(VC_a)_{ij:rj} = \beta \frac{n_{rj}}{n_{r.}} U_{ij,rj} = a'_{rj}U_{ij,rj}, \quad \text{say,}$$

$$(VC_a)_{ij:rs} = \beta \frac{n_{rj}}{n_{r.}} U_{ij,rs} = a'_{rj}U_{ij,rs}.$$

Now

$$\tfrac{1}{2}\operatorname{cov}(T_a T_b) = \operatorname{tr}(VC_a)(VB)$$
$$= \sum_i \sum_j \operatorname{tr}(VC_a \cdot VB)_{ij:ij}$$
$$= \sum_i \sum_j [\operatorname{tr} \sum_r \sum_s (VC_a)_{ij:rs}(VB)_{rs:ij}].$$

Applying the identity (9) and results (5) and (6) again, and using the forms of the elements of the sub-matrices of VC_a and VB given above, gives

$$\tfrac{1}{2}\operatorname{cov}(T_a, T_b)$$
$$= \sum_i \sum_j n_{ij}\{n_{ij}a_{ij}b_{ij} + \sum_{s \neq j} n_{is}a_{ij}b'_{ij} + \sum_{r \neq i} n_{rj}a'_{rj}b_{rj} + \sum_{r \neq i} \sum_{s \neq j} n_{rs}a'_{rj}b'_{rj}\}.$$

On substituting for the a's and b's, this reduces to

$$\tfrac{1}{2}\operatorname{cov}(T_a, T_b) = \sum_i \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}} (n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2.$$

5.9. We have now found some of the variances and covariances of T_a, T_b, T_{ab} and T_f. These and those which follow from them by symmetry, are summarized in the following table.

VARIANCES AND COVARIANCES OF UNCORRECTED SUMS OF SQUARES

var (T_a)

$$= 2\left\{\sum_i \left[\sum_j n_{ij}\,(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}\right]^2 + \sum_i \sum_{i' \neq i} \frac{(\sum_j n_{ij}n_{i'j})^2}{n_{i.}n_{i'.}}\,\beta^2\right\}$$

var (T_b)

$$= 2\left\{\sum_j \left[\sum_i n_{ij}(n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/n_{.j}\right]^2 + \sum_j \sum_{j' \neq j} \frac{(\sum_i n_{ij}n_{ij'})^2}{n_{.j}n_{.j'}}\,\alpha^2\right\}$$

var (T_{ab})

$$= 2 \sum_i \sum_j n_{ij}[n_{ij}(\alpha + \beta + \gamma + \epsilon/n_{ij})^2 + (n_{i.} - n_{ij})\alpha^2 + (n_{.j} - n_{ij})\beta^2]$$

var (T_f)

$$= 2\left[\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/N\right]^2$$

cov (T_a, T_b)

$$= 2 \sum_i \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}}\,(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2$$

cov (T_a, T_{ab})

$$= 2\left\{\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)^2/n_{i.} + \beta^2 \sum_i \sum_j n_{ij}^2(n_{.j} - n_{ij})/n_{i.}\right\}$$

cov (T_a, T_f)

$$= 2\sum_i \sum_j \frac{n_{ij}}{N}\,(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)\left(n_{i.}\alpha + \beta\,\frac{\sum_j n_{.j}n_{ij}}{n_{i.}} + \gamma\,\frac{\sum_j n_{ij}^2}{n_{i.}} + \epsilon\right)$$

cov (T_b, T_{ab})

$$= 2\left\{\sum_i \sum_j n_{ij}(n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2/n_{.j} + \alpha^2 \sum_i \sum_j n_{ij}^2(n_{i.} - n_{ij})/n_{.j}\right\}$$

cov (T_b, T_f)

$$= 2 \sum_i \sum_j \frac{n_{ij}}{N}\,(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)\left(\alpha\,\frac{\sum_i n_{i.}n_{ij}}{n_{.j}} + \beta n_{.j} + \gamma\,\frac{\sum_i n_{ij}^2}{n_{.j}} + \epsilon\right)$$

cov (T_{ab}, T_f)

$$= 2 \sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2/N$$

The expressions in the above table are those of the elements of the matrix var $(\mathbf{tt}')$ of Eq. (4). These elements are quadratic functions of the variance components α, β, γ, and ϵ, with coefficients being sums of functions of the n_{ij}'s. The other terms in (4) are not such as would simplify var ($\mathbf{vv}'$) if the elements of var ($\mathbf{tt}'$) as now known were inserted into (4), and therefore, as in any numerical case after calculating the expressions in the table these steps will be quite straightforward, it seems convenient to leave the results in their present form.

6. Balanced Data. It is easily shown that the formulae developed in the last paragraph reduce to the well-known results for balanced data when all the n_{ij} are put equal to n. For example, consider the variance of S_a. From the Analysis of Variance table, the expected value of S_a is given by

$$E(S_a) = (a - 1)(bn\alpha + n\gamma + \epsilon).$$

Then

$$\begin{aligned} \text{var}\,(T_a) &= 2[a(bn\alpha + n\beta + n\gamma + \epsilon)^2 + a(a - 1)n^2\beta^2], \\ \text{var}\,(T_f) &= 2(bn\alpha + an\beta + n\gamma + \epsilon)^2, \\ \text{cov}\,(T_a, T_f) &= 2(bn\alpha + an\beta + n\gamma + \epsilon)^2. \end{aligned}$$

Hence,

$$\begin{aligned} \text{var}\,(S_a) &= \text{var}\,(T_a - T_f) \\ &= 2(a - 1)(bn\alpha + n\gamma + \epsilon)^2 \\ &= 2[E(Sa)]^2/(a - 1) \end{aligned}$$

and with $M_a = S_a/(a - 1)$, this gives the familiar result for mean squares

$$\text{var}\,(M_a) = 2[E(M_a)]^2/(a - 1).$$

Results similar to this can be obtained for M_b and M_{ab}, the mean squares for B-effects and interaction.

7. Conclusion. Matrix methods have been developed for finding the sampling variances of estimates of components of variance. In earlier work (4) these were used for data in a 1-way classification, and this paper has extended them to data for a 2-way classification, with unequal numbers of observations in the subclasses. The estimates of the components of variance for main effects and interaction are expressed as linear functions of the corrected sums of squares and the estimate of the error variance component. By expressing the corrected sums of squares as functions of the uncorrected sums of squares, the variance-covariance matrix of the estimates of the components of variance has been expressed as a function of that for the uncorrected sums of squares, (Eq. 4). Expressions have then been found for the elements of this, the variance-covariance matrix of the uncorrected sums of squares. It has been checked that when the data are assumed balanced, i.e., all n_{ij} equal to n, these expressions reduce to the appropriate forms for variances of mean squares then having independent χ^2-distributions. Estimates with any optimum properties have not been obtained, and it would seem that the only feasible estimation procedure in a practical case would be that of replacing the variance components in these formulae by their estimates.

It is hoped that these methods can next be extended to data in a 3-way classification with unequal subclass numbers, still based on Eisenhart's Model II and using Henderson's Method I for estimation.

REFERENCES

[1] S. L. Crump, "Present status of variance components analysis," *Biometrics*, Vol. 7 (1951), pp. 1–16.

[2] C. Eisenhart, "Assumptions underlying analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1-21.

[3] C. R. Henderson, "Estimation of variance and covariance components," *Biometrics*, Vol. 9 (1953), pp. 226–252.

[4] S. R. Searle, "Matrix methods in components of variance and covariance analysis," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 737–748.

ON SOME DISTRIBUTIONS RELATED TO THE STATISTIC D_n^+

By Z. W. Birnbaum and Ronald Pyke[1]

University of Washington

1. Introduction and summary. Let $X_1 < X_2 < \cdots < X_n$ be a sample of size n, ordered increasingly, of a one-dimensional random variable X which has the continuous cumulative distribution function F. It is well known, [1], that the statistic

$$D_n^+ = \sup_{-\infty < x < +\infty} \{F_n(x) - F(x)\}, \tag{1}$$

where $F_n(x)$ is the empirical distribution function determined by $X_1, X_2, \cdots, X_n$, has a probability distribution independent of F. One may, therefore, assume that X has the uniform distribution in (0, 1) and, observing that the supremum in (1) must be attained at one of the sample points, write without loss of generality

$$D_n^+ = \max_{1 \leq i \leq n} (i/n - U_i), \tag{2}$$

where $U_1 < U_2 < \cdots < U_n$ is an ordered sample of a random variable with uniform distribution in (0, 1).

For a given $n > 0$ define the random variable i^* as that value of i, determined uniquely with probability 1, for which the maximum in (2) is reached, i.e., such that

$$D_n^+ = i^*/n - U_{i^*}, \tag{3}$$

and write

$$U_{i^*} = U^*. \tag{3.1}$$

The main object of this paper is to obtain the distribution functions of (i^*, U^*), of i^* and of U^*. The asymptotic distribution of $\alpha_n = i^*/n$ is also investigated, and bounds are obtained on the difference between the exact and the asymptotic distribution.

A number of general identities, which are not commonly known, have been verified and used in proving the above-mentioned results. Since these identities may be helpful in other problems of this type, they are separated from the main proofs and appear in the next section.

Received March 25, 1957; revised July 26, 1957.

[1] Research under the sponsorship of the Office of Naval Research. The second author's research was also supported by the Ontario Research Foundation. This paper was presented at the Seattle meeting of the Institute in August, 1956.

2. Some useful lemmas.

LEMMA 1. *For all real a, b and integer* $n \geqq 0$

$$\sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-1} = n! \sum_{i=0}^{n} \frac{(a+b)^i}{i!}. \tag{4}$$

PROOF. The identity

$$(b-n) \sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} = (a+b)^n, \tag{5}$$

for all real a, b and integer $n \geqq 0$ (for $b = n$ the left-hand term is defined as the limit for $b \to n$) was proven by Abel ([2], Vol. 1, p. 102). Denoting the left side of (4) by $f_n(a, b)$ we have

$$f_n(a, b) - nf_{n-1}(a, b) = (b-n) \sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} = (a+b)^n$$

by (5). For $n = 1$, (4) is obviously true. Assuming it is true for $n - 1$ we have

$$f_n(a, b) = nf_{n-1}(a, b) + (a+b)^n = n! \sum_{i=0}^{n} \frac{(a+b)^i}{i!},$$

which completes the proof of (4) by induction.

LEMMA 2. *For all real a, b and integers* $n \geqq 0$

$$\sum_{i=0}^{n-1} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} = \sum_{i=0}^{n-1} (a+b)^i (a+n)^{n-i-1}. \tag{6}$$

PROOF. For $b \neq n$, the left side of (6) is by Abel's identity (5) equal to

$$[(a+b)^n - (a+n)^n](b-n)^{-1},$$

which is equal to the right-hand side of (6), summed as a geometric progression. That (6) is true for $b = n$ follows from the continuity of both sides of (6).

LEMMA 3. *For all real a, b and integers* $n > 0$

$$\begin{aligned} &(a-1)(b-n) \sum_{i=0}^{n} \frac{1}{i+1} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} \\ &\qquad = \frac{1}{n+1} [(a+b)^n (a+b-n-1) - (b+1)^n (b-n)]. \end{aligned} \tag{7}$$

PROOF. Since $(a-1)/(i+1) = (a+i)/(i+1) - 1$, we may write

$$\begin{aligned} &(a-1)(b-n) \sum_{i=0}^{n} \frac{1}{i+1} \binom{n}{i} (a+i)(b-i)^{n-i-1} \\ &\qquad = \frac{b-n}{n+1} \sum_{i=0}^{n} \binom{n+1}{i+1} (a-1+i+1)^{i+1} (b+1-i-1)^{n+1-(i+1)-1} \\ &\qquad - (b-n) \sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-i-1}. \end{aligned}$$

Applying Lemma 2 to the first sum and identity (5) to the second, one concludes that this is equal to the right-hand side of (7).

COROLLARY 1. *For all integers $n > 0$ we have*

$$\sum_{i=0}^{n-1} \frac{1}{i+1}\binom{n}{i} i^i (n-i)^{n-i-1} = (n+1)^{n-1}. \tag{8}$$

PROOF. For $a = 0$, (7) yields

$$\sum_{i=0}^{n-1} \frac{1}{i+1}\binom{n}{i} i^i (b-i)^{n-i-1} = \frac{1}{n+1}\left[\frac{n^n - b^n}{n-b} - b^n + (b+1)^n\right]$$

and (8) follows for $b \to n$.

3. The distributions of (i^*, U^*), i^* and U^*. The following notations will be used: for any function f, denote by $[f \varepsilon B]$ that subset of the domain of f on which f takes values in B, a subset of the range of f; for any univariate distribution function, F, let P_F denote the n-dimensional product measure determined by the probability measure associated with F; without a subscript, P will be that measure determined by the uniform distribution function; the value of n, though suppressed in the notation, shall always be made clear by the particular circumstances of its use; furthermore, for $j = 1, 2, \cdots n$ and $u \varepsilon [0, 1]$, set

$$p_j = P[i^* = j], \qquad G^*(u, j) = P[U^* \leqq u, i^* \leqq j], \qquad H^*(u) = P[U^* \leqq u]; \tag{9}$$

for real x, $[x]$ denotes the greatest integer less than x.

All the theorems of this section are stated at the outset, and the proofs are then presented in what appears a natural sequence.

THEOREM 1. *The probabilities for i^* are given by*

$$p_j = n^{-n} \sum_{i=n-j}^{n-1} \frac{1}{i+1}\binom{n}{i} i^i (n-i)^{n-i-1} \tag{10}$$

for $j = 1, 2, \cdots n$.

THEOREM 2. *The joint probability distribution of i^* and U^* is given for all $u \varepsilon [0, 1]$ by*

$$G^*(u, k) = \sum_{j=1}^{k} K(u, j) \qquad (k = 1, 2, \cdots n),$$

where

$$K(u, j) = P[U^* \leqq u, i^* = j] =$$

$$\begin{cases} p_j & \text{if } nu \geqq j \\ n^{-1} \sum_{i=j}^{n} \binom{n}{i} (i-u)^{n-i-1}(i-nu)n^{-i} \sum_{t=[nu]}^{i} \binom{i}{t} (nu-t-1)^{i-t}(t+1)^{t-1} & \\ & \text{if } nu < j. \end{cases} \tag{11}$$

THEOREM 3. *The random variable U^* is uniformly distributed over* [0, 1].

PROOF OF THEOREM 2. For $j = 1, 2, \cdots, n$, consider the events

$$B_j = [U^* < u; i^* = j]$$
$$= [U_j < u; j/n - U_j > i/n - U_i, (i \neq j)].$$

Employing the transformation

$$Z_j = U_j; \qquad Z_i = U_i - U_j, \qquad (i \neq j)$$

one obtains

$$B_j = [Z_j \leqq u; Z_i > (i - j) / n, (i \neq j)].$$

Setting $\mathbf{Z} = (Z_1, Z_2, \cdots Z_n)$, the joint probability element of $\mathbf{Z}$ for j fixed is

$$dH_j(\mathbf{Z}) = n!\, dZ_1\, dZ_2 \cdots dZ_n$$

for

$$[-Z_j \leqq Z_1 < Z_2 \leqq \cdots \leqq Z_{j-1} \leqq 0 \leqq Z_{j+1} \leqq \cdots \leqq Z_n \leqq 1 - Z_j]$$

and zero elsewhere. Assume u and j fixed such that $nu < j$ and $1 < j < n$. Writing $\lambda = [nu]$, one has

$$\begin{aligned} K(u, j) &= \int_{B_j} dH_j(Z) \\ &= n! \int_{-1/n}^{0} dZ_{j-1} \int_{-2/n}^{Z_{j-1}} dZ_{j-2} \cdots \int_{-\lambda/n}^{Z_{j-\lambda+1}} dZ_{j-\lambda} \int_{-u}^{Z_{j-\lambda}} dZ_{j-\lambda-1} \cdots \int_{-u}^{Z_2} dZ_1 \\ &\quad \cdot \int_{-Z_1}^{u} dZ_j \int_{(n-j)/n}^{1-Z_j} dZ_n \int_{(n-j-1)/n}^{Z_n} dZ_{n-1} \cdots \int_{1/n}^{Z_{j+2}} dZ_{j+1}. \end{aligned}$$

By the linear transformation

$$x_i = \begin{cases} Z_{j+i} & \text{for } i = 1, 2, \cdots, n - j, \\ 1 - Z_j & \text{for } i = n - j + 1, \\ 1 + Z_{i-n+j-1} & \text{for } i = n - j + 2, n - j + 3, \cdots, n, \end{cases}$$

one obtains

$$\begin{aligned} K(u, j) &= n! \int_{(n-1)/n}^{1} dx_n \cdots \int_{(n-\lambda)/n}^{x_{n-\lambda+2}} dx_{n-\lambda+1} \int_{1-u}^{x_{n-\lambda+1}} dx_{n-\lambda} \cdots \\ &\quad \int_{1-u}^{x_{n-j+2}} dx_{n-j+1} \int_{(n-j)/n}^{x_{n-j+1}} dx_{n-j} \cdots \int_{2/n}^{x_3} dx_2 \int_{1/n}^{x_2} dx_1. \end{aligned} \tag{12}$$

Denote by J_k the result of integration up to and including that with respect to x_k. By repeated integration one finds

$$J_{n-j} = \frac{x_{n-j+1}^{n-j}}{(n-j)!} - \frac{x_{n-j+1}^{n-j-1}}{n(n-j-1)!}.$$

Hence

$$J_{n-j+1} = \frac{x_{n-j+2}^{n-j+1}}{(n-j+1)!} - \frac{x_{n-j+2}^{n-j}}{n(n-j)!} + W_j(u, j),$$

where for $i = 0, 1, 2, \cdots, j$,

$$W_i(u, j) = \frac{(1-u)^{n-i}}{n(n-i+1)!}(nu - i + 1).$$

Repeated integration gives

$$J_{n-\lambda} = \frac{x_{n-\lambda+1}^{n-\lambda}}{(n-\lambda)!} - \frac{x_{n-\lambda+1}^{n-\lambda-1}}{n(n-\lambda-1)!} + \sum_{i=\lambda+1}^{j} W_i(u, j)\,\frac{(x_{n-\lambda+1} - 1 + u)^{i-\lambda-1}}{(1-\lambda-1)!}.$$

By properties of the binomial expansion one obtains

$$\frac{x_{n-\lambda+1}^{n-\lambda}}{(n-\lambda)!} - \frac{x_{n-\lambda+1}^{n-\lambda-1}}{n(n-\lambda-1)!} = \frac{-1}{(n-\lambda)!}\sum_{s=0}^{n-\lambda}\binom{n-\lambda}{s}(x_{n-\lambda+1} - 1 + u)^s \cdot(1-u)^{n-\lambda-s-1}[u - (s+\lambda)/n]$$

and therefore

$$J_{n-\lambda} = \frac{1}{(n-\lambda)!}\sum_{s=j-\lambda}^{n-\lambda}\binom{n-\lambda}{s}(x_{n-\lambda+1} - 1 + u)^s \cdot(1-u)^{n-\lambda-s-1}[(s+\lambda)/n - u]. \tag{13}$$

The identity

$$\int_{(n-1)/n}^{1} dx_n \int_{(n-2)/n}^{x_n} dx_{n-1} \cdots \int_{(n-\lambda)/n}^{x_{n-\lambda+2}} (u - 1 + x_{n-\lambda+1})^s\, dx_{n-\lambda+1} \cdots dx_{n-1}\, dx_n$$
$$= \frac{s!}{(s+\lambda)!} n^{-(s+\lambda)}\left[(nu)^{s+\lambda} - \sum_{t=0}^{\lambda-1}\binom{s+\lambda}{t}(nu - 1 - t)^{s+\lambda-t}(1+t)^{t-1}\right]$$

is easily proven by induction on λ. Applying (5) one shows that the right side of this identity is equal to

$$\frac{s!\,n^{-(s+\lambda)}}{(s+\lambda)!}\sum_{t=\lambda}^{s+\lambda}\binom{s+\lambda}{t}(nu - t - 1)^{s+\lambda-t}(1+t)^{t-1}.$$

Hence it follows from (13) that

$$K(u, j) = n!\,J_n$$
$$= \frac{1}{n}\sum_{s=j-\lambda}^{n-\lambda}\binom{n}{s+\lambda}(1-u)^{n-s-\lambda-1}(s+\lambda-nu)n^{-s-\lambda}\sum_{t=\lambda}^{s+\lambda}\binom{s+\lambda}{t} \cdot(nu - t - 1)^{s+\lambda-t}(t+1)^{t-1},$$

which is the expression in (11) in the case $nu < j$.

With a few minor changes, the above argument may be also used to prove Theorem 2 for $j = 1$ and $j = n$. For example, in the discussion preceding (12) one has to define $Z_0 = Z_{n+1} = 0$, in (12) $J_0 = 1$, and in (13) $x_{n+1} = 1$. Since $j = 1$ has $\lambda = 0$, the theorem in this case follows directly from (13).

To complete the proof of Theorem 2, it remains to consider the case of $u \geqq j/n$. Since $D_n^+ \geqq 0$ then, by (3), $U^* \leqq i^*/n$. This implies that for $u > j/n$ we have

$$[U^* < u, i^* = j] = [U^* \leqq j/n, i^* = j] = [i^* = j],$$

hence the first statement of (11) is true.

Proof of Theorem 1. We have

$$\begin{aligned} p_j &= P[i^* = j] = P[U^* \leqq j/n, i^* = j] \\ &= \lim_{u \nearrow j/n} K(u, j) \\ &= \frac{1}{n} \sum_{i=j}^{n} \binom{n}{i} (1 - j/n)^{n-i-1} (i - j) n^{-i} \sum_{t=j-1}^{i} \binom{i}{t} (j - t - 1)^{i-t} (t + 1)^{t-1}, \end{aligned}$$

which, after neglecting zero terms and interchanging summations, becomes for $s = i - t$

$$\begin{aligned} p_j &= n^{-n} \sum_{t=j}^{n} \binom{n}{t} (t + 1)^{t-1} \sum_{s=0}^{n-t} \binom{n-t}{s} (j - t - 1)^s (n - j)^{n-t-s-1} (s + t - j) \\ &= n^{-n} \sum_{t=j}^{n} \binom{n}{t} (t + 1)^{t-1} (n - t - 1)^{n-t-1} \end{aligned}$$

by a direct application of the binomial expansion. Setting $i = n - t - 1$ for $t < n$, one obtains

$$p_j = n^{-n}(n + 1)^{n-1} - n^{-n} \sum_{i=0}^{n-j-1} \frac{1}{i + 1} \binom{n}{i} i^i (n - i)^{n-i-1},$$

the last sum being zero for $j = n$. By Corollary 1 it follows that for all j,

$$p_j = \frac{1}{n} \sum_{i=n-j}^{n-1} \frac{1}{i + 1} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(i - \frac{i}{n}\right)^{n-i-1}.$$

This completes the proof of Theorem 1.

Proof of Theorem 3. With $\lambda = [nu]$ as above, it follows from Theorem 2 that

$$\begin{aligned} H^*(u) &= \sum_{j=1}^{n} K(u, j) \\ &= \sum_{j=1}^{\lambda} p_j + \sum_{j=\lambda+1}^{n} \frac{1}{n} \sum_{i=j}^{n} \binom{n}{i} (1 - u)^{n-i-1} (i - nu) n^{-i} \\ &\quad \cdot \sum_{t=\lambda}^{i} \binom{i}{t} (nu - t - 1)^{i-t} (t + 1)^{t-1}. \end{aligned}$$

Interchanging summations in the last term according to the pattern

$$\sum_{j=\lambda+1}^{n} \sum_{i=j}^{n} \sum_{t=\lambda}^{i} = \sum_{i=\lambda+1}^{n} (i - \lambda) \sum_{t=\lambda}^{i} = \sum_{t=\lambda}^{n} \sum_{i=t}^{n} (i - \lambda)$$

(the second step follows because the index j does not appear in the summand; the last step follows since at $i = \lambda$ the summand is zero), one obtains

$$
\begin{aligned}
H^*(u) &= \sum_{j=1}^{\lambda} p_j + n^{-n} \sum_{t=\lambda}^{n} \binom{n}{t} (t+1)^{t-1} \\
&\quad \cdot \sum_{i=t}^{n} (i-\lambda)(i-nu) \binom{n-t}{i-t} (n-nu)^{n-i-1}(nu-t-1)^{i-t}.
\end{aligned} \tag{14}
$$

Using known properties of the binomial expansion, one can show that, whenever $n - t \neq 1$

$$
\begin{aligned}
&\sum_{s=0}^{n-t} \{(t-\lambda)(t-nu) + s(2t-\lambda-nu) + s^2\} \\
&\quad \cdot \binom{n-t}{s} (nu-t-1)^s (n-nu)^{n-t-s-1} = -(t-\lambda)(n-t-1)^{n-t-1}.
\end{aligned} \tag{15}
$$

When $n - t = 1$, this sum reduces to

$$
\begin{aligned}
&\sum_{s=0}^{1} \{(n-1-\lambda)(n-1-nu) + s(2n-2-\lambda-nu) + s^2\}(-1)^s \\
&\qquad = -(n-1-\lambda) - n(1-u).
\end{aligned} \tag{16}
$$

Substituting (15) and (16) into (14), while setting $i - t = s$, one obtains

$$
\begin{aligned}
H^*(u) &= \sum_{j=1}^{\lambda} p_j - n^{-n} \sum_{t=\lambda}^{n-1} \binom{n}{t} (t-\lambda)(t+1)^{t-1}(n-t-1)^{n-t-1} \\
&\quad - (1-u) - n^{-n}(n-\lambda)(n+1)^{n-1}.
\end{aligned} \tag{17}
$$

Employing Theorem 1, Corollary 1 with $i = n - t - 1$, and Lemma 2, one concludes from (17)

$$
\begin{aligned}
H^*(u) &= n^{-n} \sum_{i=n-\lambda}^{n-1} \frac{\lambda - n + i + 1}{i+1} \binom{n}{i} i^i (n-i)^{n-i} - n^{-n} \sum_{i=0}^{n-\lambda-1} \frac{n-i-1-\lambda}{i+1} \cdot \\
&\qquad \binom{n}{i} i^i (n-i)^{n-i-1} - (1-u) - (n-\lambda)p_n \\
&= n^{-n} \sum_{i=0}^{n-1} \binom{n}{i} i^i (n-i)^{n-i} - 1 + u \\
&= u.
\end{aligned}
$$

This completes the proof of Theorem 3.

A consequence of Theorem 1 is the following

COROLLARY 2. *For all integers* $n > 0, j > 0$,

$$
0 < p_1 < p_2 < \cdots < p_n < 1,
$$

$$
\lim_{n\to\infty} np_j = \sum_{i=1}^{j} \frac{e^{-i} i^{i-1}}{i!},
$$

$$
\lim_{n\to\infty} np_{n-j} = e - \sum_{i=0}^{j-1} \frac{e^{-i} i^i}{(i+1)!}.
$$

PROOF. The first statement is evident from (10); the second follows from (10) by applying Stirling's formula; and the third follows by applying Stirling's formula to the expression

$$p_{n-j} = \frac{(n+1)^{n-1}}{n^n} - \sum_{i=0}^{j-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1},$$

which can be obtained from (10) by Corollary 1.

Thus the statistic D_n^+ places more weight upon the larger observations than on the smaller ones, in the sense that the maximum deviation between F and F_n is more probable to occur at X_{k+1} than at X_k for $k = 1, 2, \cdots, n-1$.

4. The asymptotic distribution of $\alpha_n = i^*/n$. Writing U_n^* instead of U^*, we have according to Theorem 3,

$$P[U_n^* \leqq u] = H_n^*(u) = u, \qquad 0 \leqq u \leqq 1. \tag{18}$$

Since the Glivenko-Cantelli theorem ([3], p. 260) implies that D_n^+ converges in probability to zero, it follows from (3) that

$$\alpha_n - U_n^* \rightarrow 0 \quad \text{in probability.} \tag{19}$$

From (18) and (19) one can conclude that α_n is asymptotically uniformly distributed on [0, 1].

The following theorem contains more specific statements on the asymptotic behavior of the distribution of α_n.

THEOREM 4. *For every positive integer n we have*

$$E(\alpha_n) = \frac{1}{2} + \frac{1}{2} n^{-n-1} n! \sum_{i=0}^{n-1} \frac{n^i}{i!}, \tag{20}$$

$$x - \sqrt{n^{-n-1} n! \sum_{i=0}^{n-1} \frac{n^i}{i!}} \leqq \Pr\{\alpha_n < x\} \leqq x \quad \text{for } 0 \leqq x \leqq 1. \tag{21}$$

PROOF OF THEOREM 4. From Theorem 1 we have

$$\begin{aligned} E(\alpha_n) &= n^{-n-1} \sum_{j=1}^{n} \sum_{i=n-j}^{n-1} \frac{j}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} \\ &= \frac{1}{2}\left(1 - \frac{1}{n}\right) + \frac{1}{2} n^{-n-1} \sum_{i=0}^{n} \binom{n}{i} i^i (n-i)^{n-i} \end{aligned}$$

and this by Lemma 1 yields (20). To obtain the upper bound on $\Pr\{\alpha_n < x\}$ in (21) we note that

$$G_n(x) = \Pr\{\alpha_n < x\} = \sum_{u=1}^{[nx]} p_j$$

and in view of Corollary 2 this must be $<x$ for all $1/n < x \leqq 1$.

To obtain the lower inequality in (21) we need

LEMMA 4. *Let X be a random variable with c.d.f. F, such that $F(0) = 0$, $F(1 + 0) = 1$, $F(x) \leqq x$ for $0 \leqq x \leqq 1$. Then*

$$F(s) \geqq s - \sqrt{2E(X) - 1}. \tag{22}$$

PROOF OF LEMMA 4. We have

$$E(X) = \int_0^1 X\,dF(X) \geqq 1 - \int_0^1 F(X)\,dX$$

$$\geqq 1 - \int_0^{F(s)} X\,dX - \int_{F(s)}^{s} F(s)\,dX - \int_s^1 X\,dX = \tfrac{1}{2}\{1 + [s - F(s)]^2\}$$

and this implies (22). One verifies directly that, for given s and $F(s)$, equality is attained in (22) when $F(t) = t$ for $0 \leqq t \leqq F(s)$, $F(t) = F(s)$ for $F(s) \leqq t \leqq s$, $F(t) = t$ for $s \leqq t \leqq 1$.

According to the upper inequality in (21), Pr $\{\alpha_n < x\}$ fulfills the assumptions of Lemma 4, which together with (20) yields the lower bound of (21).

It may be noted that by an application of Stirling's formula one obtains from (21)

$$0 \leqq x - \Pr\{\alpha_n < x\} = 0(n^{-1/4}), \tag{23}$$

and that (20) together with (3) yields

$$E(D_n^+) = 2^{-1}n^{-n-1}n! \sum_{i=0}^{n-1} \frac{n^i}{i!}. \tag{24}$$

REFERENCES

[1] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[2] N. H. ABEL, *Oevres Complètes*, Christiania, C. Groendahl, 1839.

[3] M. FRÉCHET, *Recherches théoriques modernes sur la théorie des probabilités*, Gauthier-Villars, Paris, 1937, p. 260.

ON SEVERAL STATISTICS RELATED TO EMPIRICAL DISTRIBUTION FUNCTIONS[1]

By Meyer Dwass

Northwestern University

1. Introduction. Let $X_1, \cdots, X_n$ be n independent random variables, each with the same continuous c.d.f., $F(x)$. Let $F_n(x)$ be the empirical c.d.f. of the X_i's. We consider the following random variables,

$$U_n = \mu\{F(t):F_n(t) - F(t) > 0\},$$

$$D_n = \sup_{-\infty<t<\infty} (F_n(t) - F(t)),$$

$$V_n = \inf_{-\infty<t<\infty} \{F(t):F_n(t) - F(t) = D_n\},$$

where $\{F(t): \quad \}$ denotes the set of values of $F(t)$, for which t satisfies the condition after the colon. These are sets in the interval (0, 1). In the definition of U_n, $\mu\{ \quad \}$ means Lebesgue measure. Obviously, there is no loss of generality in supposing that the X_i's are uniformly distributed over (0, 1) and hence

$$\text{(1)} \qquad \begin{cases} U_n = \mu\{t:F_n(t) - t > 0\}, \\ D_n = \sup\limits_{0\leq t\leq 1} (F_n(t) - t), \\ V_n = \inf\limits_{-\infty<t<\infty} \{t:F_n(t) - t = D_r\}. \end{cases}$$

In [5], Kac showed that as $n \to \infty$, U_n has an asymptotic distribution which is uniform over (0, 1). A stronger result was recently obtained by Gnedenko and Mihalevič in [4] in which they showed that *for every n, U_n is uniformly distributed*. Birnbaum and Pyke in a forthcoming paper [2] show that for every n, V_n is also distributed uniformly over (0.1). The methods of [2] and [4] are computational and the purpose of this note is to derive the uniform distribution of U_n and V_n by a short method which employs results of E. S. Andersen and a well-known relationship between the Poisson process and uniformly distributed random variables. In Sec. 3, a generalization of these results is given.

2. Proof of uniform distribution of U_n and V_n. The proof depends on two sets of facts. The first refers to the Poisson process. By this we mean the stochastic process, $X(t)$, with independent and stationary Poisson distributed increments, defined for $t \geqq 0$ and such that $X(0) = 0$. For this process, it is well known that given that $X(1) = n$, a positive integer, then the conditional distribution of the discontinuity (jump) points, $t_1 \leqq t_2 \leqq \cdots \leqq t_n$ of $X(t)$, $0 \leqq t \leqq 1$, is that

Received February 11, 1957; revised June 18, 1957.

[1] Sponsored by Air Force Office of Scientific Research AFOSR-TN-57-784, AD148015 Contract No. AF 49(638)-151.

of the ordered values of n independent, uniform random variables. Another way of saying this, somewhat roughly, is that the conditional distribution of the random function $X(t)$, $0 \leqq t \leqq 1$, given that $X(1) = n$, is that of the empirical c.d.f. of n independent, uniform random variables. For a proof of these facts see p. 400 of [3]. The second set of needed facts is contained in a paper of E. S. Andersen [1], namely:

LEMMA (ANDERSEN). *Let $Y_1, Y_2, \cdots$ be independent and identically distributed random variables. Make the definitions*

$$S_0 = 0 \text{ (a.s.)}, \qquad S_i = \sum_{j=1}^{i} X_j,$$

$$L_r = \textit{smallest } i \textit{ for which } S_i = \max(0, S_1, \cdots, S_r).$$

$$N_r = \textit{number of positive terms in } S_1, \cdots, S_r.$$

Then

$$P(L_r = m \mid S_{r+1} = 0) = P(N_r = m \mid S_{r+1} = 0) = \frac{1}{r+1}, \tag{2}$$

for $m = 0, 1, \cdots, r$ if and only if

$$P(S_i = S_{r+1} = 0) = 0, \qquad (i = 1, 2, \cdots, r). \tag{3}$$

We remark that Andersen's results are much more general, but we state them in a form convenient for our applications.

THEOREM 1. U_n *and* V_n *are each distributed uniformly over* $(0, 1)$.

PROOF. Consider the Poisson process $X(t)$, $0 \leqq t \leqq 1$. Divide the interval $(0, 1)$ into the $r + 1$ parts $(0, 1/(r + 1))$, $(1/(r + 1), 2/(r + 1))$, $\cdots$, $(r/(r + 1), 1)$, where $r + 1$ is greater than n and *is a prime number*. (Whenever we state $r \to \infty$ we will understand that $r + 1$ goes through the primes.) The increments of $X(t)$ in these intervals are independent and identically distributed Poisson random variables. We denote these increments by $W_1, W_2, \cdots, W_{r+1}$, respectively, and define $Y_i = W_i - n/(r + 1)$, $i = 1, \cdots, r + 1$. The Y_i's are independent and identically distributed. We want to show that they satisfy (3) of Andersen's lemma. This is so because $S_i = S_{r+1} = 0$ implies that $(r + 1) \cdot X(i/(r + 1)) = ni$. This cannot hold since by the primeness of $r + 1$, n must be a factor of $X(i/r + 1)$, but since $X(t)$ is non-decreasing this would mean $X(i/(r + 1)) = n$, or $r + 1 = i$, a contradiction; thus (3) holds. Under the condition $X(1) = n$, $X(t)$ is distributed like $F_n(t)$, for $s \leqq t \leqq 1$. Hence we can define U_n, V_n for $X(t)$, $0 \leqq t \leqq 1$. We next observe that when $X(1) = n$, then

$$\left| U_n - \frac{N_r}{r+1} \right| < \frac{A}{r+1}, \qquad \left| V_n - \frac{L_r}{r+1} \right| < \frac{B}{r+1}, \tag{4}$$

where A, B are constants which depend on n but not on r. Thus, under the condition $X(1) = n$, both absolute values in (4) converge in probability to zero as $r \to \infty$. Since $N_r/(r + 1)$ and $L_r/(r + 1)$ are asymptotically uniformly distributed over $(0, 1)$ as $r \to \infty$, this completes the proof.

3. Generalization. A generalization of Theorem 1 can be given which may be of interest. Let

$$X_{11}, \cdots, X_{1n_1}; \cdots; X_{k1}, \cdots, X_{kn_k},$$

be $n = n_1 + \cdots + n_k$ independent random variables each uniformly distributed over $(0, 1)$. Let $F^{(1)}(t), \cdots, F^{(k)}(t)$ be the empirical c.d.f.'s of each of the k sets of variables and define

$$F_\rho(t) = \rho_1 F^{(1)}(t) + \cdots + \rho_k F^{(k)}(t), \qquad 0 \leqq t \leqq 1,$$

where $\rho = (\rho_1, \rho_2, \cdots, \rho_k)$, $\rho_i > 0$, $\rho_1 + \rho_2 + \cdots + \rho_k = 1$. In the special case where $\rho_i = n_i/n$, $i = 1, \cdots, k$, then $F_\rho(t)$ is the empirical c.d.f. of the combined set of n variables. Otherwise $F_\rho(t)$ can only be described as a nondecreasing random step function on $(0, 1)$ such that $F_n(0) = 0$, $F_n(1) = 1$. Nevertheless, random variables U_ρ, D_ρ and V_ρ analogous to U_n, D_n and V_n may be defined for $F_\rho(t)$ exactly as was done in (1) for $F_n(t)$; (replace $F_n(t)$ by $F_\rho(t)$ in (1)). In the following theorem we understand them to be so defined.

Theorem 2. *U_ρ and V_ρ are each distributed uniformly over* $(0, 1)$.

Proof. Let $X_1(t), X_2(t), \cdots, X_k(t)$ be k independent Poisson processes and define $X(t) = \rho_1 X_1(t) + \cdots + \rho_k X_k(t)$. Then $X(t)$ is also a process with stationary independent increments. Define now $\rho = (\rho_1, \rho_2, \cdots, \rho_k)$,

$$\begin{cases} \tilde{U}_\rho = \mu\{t: X(t) - X(1)t > 0, 0 \leqq t \leqq 1\}, \\ \tilde{D}_\rho = \sup\limits_{0 \leqq t \leqq 1} (X(t) - X(1)t), \\ \tilde{V}_\rho = \inf\limits_{0 \leqq t \leqq 1} \{t: X(t) - X(1)t = \tilde{D}_\rho\}. \end{cases}$$

We suppose first that

$$\rho_1 = a_1/a, \cdots, \rho_k = a_k/a, \tag{5}$$

where $a_1, \cdots, a_k$ are positive integers, and $a_1 + \cdots + a_k = a$. If b is a number such that $P(X(1) = b) > 0$, then $\tilde{U}_\rho$, $\tilde{V}_\rho$ are each uniformly distributed over $(0, 1)$ given that $X(1) = b$. The proof of this fact follows exactly the proof of theorem 1. In particular the definition of the ρ_i's by (5) allows a verification of the condition (3) of Andersen's lemma which is exactly analogous to that done in the proof of Theorem 1. Since the ρ_i's as defined by (5) are dense in the set of all possible ρ_i's, it follows by a simple continuity argument that the conditional distribution of $\tilde{U}_\rho$, $\tilde{V}_\rho$ given that $X(1) = b$, is uniform *without* the restriction (5). If $X(1) = \rho_1 X_1(1) + \cdots + \rho_k X_k(1) = b$, this need not uniquely determine the values of the $X_i(1)$. That is, there may be two different sets of positive or zero integers, $x_1, \cdots, x_k$; $y_1, \cdots, y_k$, such that

$$\rho_1 x_1 + \cdots + \rho_k x_k = \rho_1 y_1 + \cdots + \rho_k y_k = b.$$

On the other hand, there is a dense subset of the k-dimensional unit cube where this cannot happen, namely any dense subset, each point of which has rationally independent coordinates. Thus, in such a dense subset $X(1) = \rho_1 n_1 + \cdots + \rho_k n_k$ if and only if $x_i(1) = n_1, \cdots, x_k(1) = n_k$, for a set of ρ_i's which are dense in the set of all possible ρ_i's. For such ρ_i's the conditional distribution of $\tilde{U}_\rho$ and $\tilde{V}_\rho$ given that $X_1(1) = n_1, \cdots, X_k(1) = n_k$, is thus uniform. This holds also for the exceptional ρ_i's by a continuity argument. This completes the proof since $F^{(1)}(t), \cdots, F^{(k)}(t)$ are distributed like $X_1(t), \cdots, X_k(t)$ for $0 \leqq t \leqq 1$, under the conditions that $X_1(1) = n_1, \cdots, X_k(t) = n_k$.

4. Concluding remarks. The linear combinations of Theorem 2 are convex ($\rho_1 + \cdots + \rho_k = 1$) and positive ($\rho_i > 0$). The convexity, as well as the strict positivity, is a matter of convenience. The condition of non-negativeness, however, cannot be removed. It is easy to verify directly, for example, that the theorem does not hold for

$$F_\rho(t) = \rho_1 F^{(1)}(t) + \rho_2 F^{(2)}(t),$$

if $\rho_1 > 0$ and $\rho_2 < 0$. The trouble arises because the condition (3) of Andersen's lemma fails to hold.

REFERENCES

[1] E. S. Andersen, "On the fluctuations of sums of random variables," *Math. Scand.*, Vol. 1 (1953), pp. 263–285.

[2] Z. W. Birnbaum and R. Pyke, "On some distributions related to the statistic D_n^+," Technical report no. 23, July 17, 1956, Laboratory of Statistical Research, Dept. of Math., University of Washington, Seattle, Washington.

[3] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, (1953).

[4] B. V. Gnedenko and V. S. Mihalevič, "Two theorems on the behavior of empirical distribution functions," *Doklady Akad. Nauk SSSR* (N.S.), Vol. 85 (1952), pp. 25–27.

[5] M. Kac, "On deviations between theoretical and empirical distributions," *Proc. Nat Acad. Sci., U.S.A.*, Vol. 35 (1949), pp. 252–257.

ON THE EFFICIENCY OF ESTIMATES OF TREND IN THE ORNSTEIN UHLENBECK PROCESS

By Charlotte T. Striebel

University of California, Berkeley

1. Summary. The problem is that of estimating the trend of a normal process when the trend function is known up to a finite number of coefficients. That is,

$$y_t = x_t + f(t), \qquad 0 \leqq t \leqq T,$$

where x_t is a normal process with mean zero and covariance function

$$E[X_u, X_v] = C(u, v)$$

and

$$f(t) = k_1\phi_1(t) + \cdots + k_s\phi_s(t).$$

The $\phi_i(t)$ are known functions and the k_i are to be estimated.

The standard procedure in such a case is to derive the estimates by the maximum likelihood method. However, if the covariance function $C(u, v)$ is not completely known, this is usually impossible, and it is essential to find an alternative procedure. The method of least squares has been proposed by Mann [1]. The estimates obtained by this method are independent of $C(u, v)$ and have the additional advantage of being easily computed. Mann and Moranda [2] showed that for the Ornstein Uhlenbeck process the asymptotic efficiency of the least square estimate relative to the maximum likelihood estimate is one, in the special case that the $\phi_i(t)$ are polynomials or trigonometric polynomials. Mann defines the efficiency $\bar{e}(T)$ of an estimate $\tilde{f}(t)$

$$\bar{e}(T) = \frac{E\left[\int_0^T [\hat{f}(t) - f(t)]^2\, dt\right]}{E\left[\int_0^T [\tilde{f}(t) - f(t)]^2\, dt\right]},$$

where $\hat{f}(t)$ is the maximum likelihood estimate. For the cases that shall be of particular interest—the Ornstein Uhlenbeck process with $\tilde{f}(t)$ a linear unbiased estimate—Mann and Moranda [2] have shown that $\bar{e}(T) \leqq 1$.

In the present paper the asymptotic efficiency of the least square estimates will be computed for a wider class of functions $\phi_i(t)$. It will be shown that except for a special case just slightly broader than the one treated by Mann and Moranda, the asymptotic efficiency is actually less than one. Thus except for this special case, the least square estimates could be improved upon. An alternative estimate $\bar{k}_i(\alpha)$ is proposed. It will be shown that for $\alpha \geqq \beta$, where β is the true correlation parameter in the Ornstein Uhlenbeck process, the estimates $\bar{k}_i(\alpha)$ are

Received April 19, 1957; revised October 31, 1957.

asymptotically more efficient than the least square estimates, and in fact as $\alpha \to \beta$ from above the efficiency increases (strictly) to one.

2. Introduction. The least square estimate is obtained by minimizing the expression

$$\int_0^T (y_t - f(t))^2\, dt$$

and is given by

$$\bar{k}_i = \sum_{j=1}^{s} G^{ij}(T) \int_0^T \phi_i(t) y_t\, dt,$$

where

$$G_{ij}(T) = \int_0^T \phi_i(t)\phi_j(t)\, dt. \tag{1}$$

The maximum likelihood estimates $\hat{k}_i$ minimize

$$\int_0^T \int_0^T [y_u - f(u)][y_v - f(v)] C^{-1}(u, v)\, du\, dv$$

and are given by

$$\hat{k}_i = \sum_{j=1}^{s} \Phi^{ij}(T) \int_0^T \int_0^T \phi_j(u) y_v C^{-1}(u, v)\, du\, dv,$$

where

$$\Phi_{ij}(T) = \int_0^T \int_0^T \phi_i(u)\phi_j(v) C^{-1}(u, v)\, du\, dv. \tag{2}$$

It will be assumed that the $\phi_i(t)$ and $C(u, v)$ are such that these integrals exist. The efficiency of the least square estimates can now be computed.

$$\bar{e}(T) = \frac{t[G(T)\Phi^{-1}(T)]}{t[\Psi(T)G^{-1}(T)]},$$

where

$$\Psi(T) = \int_0^T \int_0^T \phi_i(u)\phi_j(v) C(u, v)\, du\, dv. \tag{3}$$

The trace of the matrix is t.

It will further be assumed that there are functions $H_i(T)$ such that the limits

$$\begin{aligned} &\lim_{T\to\infty} \frac{G_{ij}(T)}{H_i(T)H_j(T)} = G_{ij}, \\ &\lim_{T\to\infty} \frac{\Phi_{ij}(T)}{H_i(T)H_j(T)} = \Phi_{ij}, \\ &\lim_{T\to\infty} \frac{\Psi_{ij}(T)}{H_i(T)H_j(T)} = \Psi_{ij} \end{aligned} \tag{4}$$

exist and are positive definite matrices. The asymptotic efficiency then is

$$\bar{e} = \lim_{T\to\infty} \bar{e}(T) = \frac{t(G\Phi^{-1})}{t(\Psi G^{-1})}. \tag{5}$$

Necessary and sufficient conditions that $\bar{e} = 1$ will be found for two classes of G, Φ, Ψ. The first, which includes the cases treated by Mann and Moranda, requires that G, Φ, Ψ be of the form

$$\begin{aligned} G &= \sum_{n=1}^{N} G_n, \\ \Phi &= \sum_{n=1}^{N} c_n G_n, \\ \Psi &= \sum_{n=1}^{N} \frac{1}{c_n} G_n, \end{aligned} \tag{6}$$

where the G_n are positive semi-definite matrices and the c_n are distinct positive real numbers. The second requires that

$$\begin{aligned} \Phi &= BGB^T \\ \Psi &= B^{-1}GB^{-1T} + C, \end{aligned} \tag{7}$$

where B is positive definite, and C is positive semi-definite.

Results will be applied to the case that x_t is an Ornstein Uhlenbeck process and the $\phi_i(t)$ are of the form first

$$\phi_i(t) = \sum_{n=1}^{N} \sum_{r=1}^{\gamma_i} t^r(a_{inr} \sin \omega_n t + b_{inr} \cos \omega_n t)$$

and second

$$\phi_i(t) = e^{a_i t}, \qquad a_i > 0.$$

When the covariance $C(u, v)$ involves some unknown parameters an attempt can be made to estimate them along with the k_i by the maximum likelihood method. However, this frequently leads to equations which cannot be solved. In this case, a natural procedure is to make an estimate $C^*(u, v)$ of $C(u, v)$ by any convenient method and then use the maximum likelihood estimates of the k_i based on the covariance $C^*(u, v)$.

For the Ornstein Uhlenbeck process

$$C(u, v) = \sigma^2 e^{-\beta|u-v|}.$$

Let

$$C^*(u, v) = \begin{cases} \sigma^2 & \text{if } u = v, \\ 0 & \text{otherwise.} \end{cases}$$

This covariance function yields the least square estimates. If the true value β is replaced by α, a family of estimates is obtained by this method.

$$\bar{k}_i(\alpha) = \sum_{j=1}^{s} \Phi_\alpha^{ij}(T) \frac{1}{2}\left[\phi_j(T)y_T + \phi_j(0)y_0 + \frac{1}{\alpha}\int_0^T \phi_j'(t)\,dy_t + \alpha\int_0^T \phi_j(t)y_t\,dt\right], \tag{8}$$

where

$$\Phi_{\alpha ij}(T) = \frac{1}{2}\left[\phi_i(T)\phi_j(T) + \phi_i(0)\phi_j(0) + \frac{1}{\alpha}\int_0^T \phi_i'(t)\phi_j'(t)\,dt + \alpha\int_0^T \phi_i(t)\phi_j(t)\,dt\right]. \tag{9}$$

Clearly

$$\lim_{\alpha\to\infty} \bar{k}_i(\alpha) = \bar{k}_i,$$

$$\bar{k}_i(\beta) = \hat{k}_i.$$

3. Efficiency of estimates for G, Φ, Ψ, **of form** (6). Assume that G, Φ, Ψ, defined by (1), (2), (3), and (4) are the special form (6). Then from (5)

$$1 - \bar{e} = \frac{t(\Psi G^{-1} - G\Phi^{-1})}{t(\Psi G^{-1})} = \frac{t[G^{-1}\Phi^{-1}(\Phi\Psi - GG)]}{t(\Psi G^{-1})},$$

$$\begin{aligned}(\Phi\Psi - GG) + (\Phi\Psi - GG)^T &= \sum_{n=1}^{N}\sum_{m=1}^{N}\left[\frac{c_n}{c_m} G_nG_m - G_nG_m + \frac{c_m}{c_n} G_nG_m - G_nG_m\right] \\ &= \sum_{n=1}^{N}\sum_{m=1}^{N} \frac{(c_m - c_n)^2}{c_nc_m} G_nG_m.\end{aligned} \tag{10}$$

G_mG_n are positive semi-definite and

$$\frac{(c_m - c_n)^2}{c_nc_m} > 0, \qquad \text{all} \quad m \neq n.$$

Thus, $\Phi\Psi - GG$ is positive semi-definite. In order that $1 - \bar{e} = 0$, it is necessary and sufficient that

$$\Phi\Psi - GG = 0.$$

This is equivalent to requiring

$$t(G_mG_n) = 0, \qquad \text{all } m \neq n.$$

This result will be stated as a theorem.

THEOREM 1. *If G, Φ, Ψ are nonsingular and of the form*

$$G = \sum_{n=1}^{N} G_n ,$$

$$\Phi = \sum_{n=1}^{N} c_n G_n ,$$

$$\Psi = \sum_{n=1}^{N} \frac{1}{c_n} G_n ,$$

where the G_n are positive semi-definite matrices and the c_n are distinct positive real numbers, then

$$\bar{e} = \frac{t(G\Phi^{-1})}{t(\Psi G^{-1})} = 1$$

if and only if

$$t(G_m G_n) = 0, \qquad \text{all } m \neq n.$$

For the Ornstein Uhlenbeck process the theorem can be applied to obtain the special result.

THEOREM 2. *Let*

$$y_t = x_t + f(t),$$

where x_t is an Ornstein Uhlenbeck process with mean zero, and

$$f(t) = k_1\phi_1(t) + \cdots + k_s\phi_s(t).$$

Suppose

$$\phi_i(t) = \sum_{n=1}^{N} \sum_{r=1}^{\gamma_i} t^r(a_{inr} \sin \omega_n t + b_{inr} \cos \omega_n t)$$

are such that

$$\phi_i^*(t) = t^{\gamma_i} \sum_{n=1}^{N} (a_{in\gamma_i} \sin \omega_n t + b_{in\gamma_i} \cos \omega_n t)$$

are linearly independent. Then the asymptotic efficiency of the least square estimates of the k_j is one, if and only if

$$\begin{aligned} \sum a_{in\gamma}a_{im\gamma} &= 0, \\ \sum a_{in\gamma}b_{im\gamma} &= 0, \qquad (11) \\ \sum b_{in\gamma}b_{im\gamma} &= 0, \end{aligned}$$

for all γ and $m \neq n$. The sums extend over all i for which $\gamma_i = \gamma$.

PROOF. Let $H_i(T) = T^{\gamma_i+1/2}$. The only terms which appear in the limits (4) will be those of maximum order, that is, those of the $\phi^*(t)$. Denote $a_{in\gamma_i}$ by a_{in} and $b_{in\gamma_i}$ by b_{in}. Then G, Φ, Ψ can be computed and are of the form (6) with

$$G_{nij} = \frac{a_{in}\, a_{jn} + b_{in}\, b_{jn}}{(\gamma_i + \gamma_j + 1)\gamma_i\,!\gamma_j\,!},$$

$$c_n = \frac{\beta^2 + \omega_n^2}{2\beta}.$$

The G_n can easily be shown to be positive semi-definite. Thus by theorem 1, $\bar{e} = 1$ if and only if

$$t(G_m G_n) = 0, \qquad \text{all } m \neq n.$$

$$t(G_m G_n) = \sum_{i=1}^{s}\sum_{j=1}^{s} \frac{a_{in}\, a_{im}\, a_{jn}\, a_{jm} + a_{in}\, b_{im}\, a_{jn}\, b_{jm} + a_{im}\, b_{in}\, a_{jm}\, b_{jn} + b_{in}\, b_{im}\, b_{jn}\, b_{jm}}{(\gamma_i + \gamma_j + 1)^2\,(\gamma_i\,!)^2\,(\gamma_j\,!)^2}$$

$$= \sum_{\gamma}\sum_{\delta} \frac{A_{mn\gamma}\, A_{mn\delta} + C_{mn\gamma}\, C_{mn\delta} + C_{nm\gamma}\, C_{nm\delta} + B_{mn\gamma}\, B_{mn\delta}}{(\gamma + \delta + 1)^2(\gamma!)^2(\delta!)^2}.$$

γ and δ are summed over all distinct values of γ_i and

$$A_{mn\gamma} = \sum a_{in} a_{im},$$

$$B_{mn\gamma} = \sum b_{in} b_{im},$$

$$C_{mn\gamma} = \sum a_{in} b_{im}.$$

The summations extend over all values of i for which $\gamma_i = \gamma$. Since

$$\frac{1}{(\gamma + \delta + 1)^2}$$

is a positive definite matrix for γ, δ ranging over distinct integers,

$$t(G_m G_n) = 0, \qquad \text{all } m \neq n$$

if and only if

$$A_{mn\gamma} = B_{mn\gamma} = C_{mn\gamma} = 0$$

for all γ and $m \neq n$.

Thus, unless the special conditions (11) are satisfied, $\bar{e}$ will be strictly less than one. For example,

$$f(t) = k_1 + k_2 \sin t + k_3 \sin 2t$$

can be estimated efficiently by least squares, but

$$f(t) = k_1 + k_2 (\sin t + \sin 2t)$$

cannot.

Grenander and Rosenblatt [3] in Section 7.6 obtain results very similar to those of Theorem 2.

THEOREM 3. *If y_t and $\phi_i(t)$ are as in the hypothesis of theorem 2, then the asymptotic efficiency $\bar{e}(\alpha)$ of the estimate $\bar{k}_i(\alpha)$ (8) is monotone decreasing from 1 at $\alpha = \beta$ to $\bar{e}$ as $\alpha \to \infty$. If $\bar{e} \neq 1$, then it is strictly decreasing.*

PROOF. First the efficiency of the $\bar{k}_i(\alpha)$ estimates must be computed.

$$\begin{aligned}
E(T, \alpha) &= E\left[\int_0^T (\hat{f}_\alpha(t) - f(t))^2\, dt\right] \\
&= t[\Sigma_{\bar{k}_i(\alpha)\bar{k}_j(\alpha)}(T) G_{ij}(T)]. \\
\Sigma_{\bar{k}_i(\alpha)\bar{k}_j(\alpha)} &= E[(\bar{k}_i(\alpha) - k_i)(\bar{k}_j(\alpha) - k_j)] \\
&= \Phi_\alpha^{-1}(T)\left[\frac{(\alpha^2 - \beta^2)}{4} \Psi_\alpha(T) + \frac{\beta}{\alpha} \Phi_\alpha(T)\right] \Phi_\alpha^{-1}(T). \\
\Psi_{\alpha ij}(T) &= \int_0^T \int_0^T \left(\phi_i(u) + \frac{1}{\alpha} \phi_i'(u)\right)\left(\phi_j(v) + \frac{1}{\alpha} \phi_j'(v)\right) e^{-\beta|u-v|}\, du\, dv,
\end{aligned}$$

and $\Phi_\alpha(T)$ is defined by (9).

For $\phi_i(t)$ as in theorem 2 and

$$H_i(T) = T^{\gamma_i + 1/2},$$

$$\Phi_{\alpha ij} = \lim_{T\to\infty} \frac{\Phi_{\alpha ij}(T)}{H_i(T) H_j(T)} = \frac{1}{2\alpha} [\alpha^2 G_{ij} + \zeta_{ij}],$$

and

$$\Phi = \frac{1}{2\beta} (\beta^2 G + \zeta),$$

where

$$\zeta_{ij} = \lim_{T\to\infty} \frac{\int_0^T \phi_i'(t)\phi_j'(t)\, dt}{H_i(T) H_j(T)}.$$

$$\Psi_{\alpha ij} = \lim_{T\to\infty} \frac{\Psi_{\alpha ij}(T)}{H_i(T) H_j(T)} = \frac{(\alpha^2 - \beta^2)}{\alpha^2} \Psi_{ij} + \frac{2\beta}{\alpha^2} G_{ij}$$

Thus,

$$\Phi_\alpha = \frac{1}{2\alpha} [(\alpha^2 - \beta^2) G + 2\beta\Phi].$$

Let

$$A = (\alpha^2 - \beta^2) G + 2\beta\Phi.$$

Then

$$\begin{aligned}
E(\alpha) &= \lim_{T\to\infty} E(T, \alpha) \\
&= t\{A^{-1}A^{-1}[(\alpha^2 - \beta^2)^2 \Psi + 4\beta(\alpha^2 - \beta^2) G + 4\beta^2 \Phi]\},
\end{aligned}$$

and

$$\frac{\partial E(\alpha)}{\partial \alpha} = 8\alpha\beta(\alpha^2 - \beta^2)t[A^{-1}A^{-1}GA^{-1}\Phi(G^{-1}\Psi - \Phi^{-1}G)].$$

It follows from (10) that

$$G^{-1}\Psi - \Phi^{-1}G$$

is positive semi-definite. Thus, for $\alpha > \beta$, the derivative of $E(\alpha)$ is nonnegative and $E(\alpha)$ is monotone increasing. If $\bar{e} \neq 1$, then

$$t(G^{-1}\Psi - \Phi^{-1}G) > 0,$$

and at least one characteristic root must be nonzero. Since $A^{-1}A^{-1}GA^{-1}\Phi$ is positive definite $\partial E(\alpha)/\partial\alpha$ will then be positive, and $E(\alpha)$ is strictly increasing.

$$\bar{e}(\alpha) = \lim_{T\to\infty} \frac{E\left[\int_0^T (\hat{f}(t) - f(t))^2\,dt\right]}{E\left[\int_0^T (\hat{f}_\alpha(t) - f(t))^2\,dt\right]}.$$

$$= \frac{t(\Phi^{-1}G)}{E(\alpha)}.$$

4. Efficiency of estimates for exponential $\phi_i(t)$. Assume G, Φ, Ψ are of the form (7). Then

$$1 - \bar{e} = \frac{t(\Psi G^{-1} - \Phi^{-1}G)}{t(\Psi G^{-1})} = \frac{t(CG^{-1})}{t(\Psi G^{-1})},$$

and $\bar{e} = 1$ if and only if $C = 0$.

Let

$$\phi_i(t) = e^{a_i t}$$

and

$$H_i(\alpha) = e^{a_i T},$$

where the a_i are positive and distinct. Then

$$G_{ij} = \frac{1}{a_i + a_j},$$

$$\Phi_{ij} = \frac{(\beta + a_i)(\beta + a_j)}{a_i + a_j},$$

$$\Psi_{ij} = \frac{2\beta}{(\beta + a_i)(\beta + a_j)(a_i + a_j)} + \frac{1}{(\beta + a_i)(\beta + a_j)}.$$

Thus,

$$B_{ij} = \frac{(\beta + a_i)\delta_{ij}}{\sqrt{2\beta}},$$

$$C = \frac{1}{2\beta}B^{-1}\,B \neq 0,$$

and hence

$$\bar{e} < 1.$$

Since the least square estimates are not asymptotically efficient, it is of interest to compute the efficiency of the $\bar{k}_i(\alpha)$ estimates. In this case

$$\Phi_{\alpha ij} = \frac{(\alpha + a_i)(\alpha + a_j)}{2\alpha(a_i + a_j)} = \frac{1}{2\alpha} AGA,$$

where

$$A_{ij} = (\alpha + a_i)\delta_{ij}$$

$$\Psi_\alpha = \frac{1}{\alpha^2} A\Psi A$$

$$E(\alpha) = t\{G^{-1}A^{-1}GA^{-1}G^{-1}[(\alpha^2 - \beta^2)\Psi + 2\beta G]\}$$

$$\frac{\partial E(\alpha)}{\partial \alpha} = 2(\alpha - \beta)t[A^{-1}G^{-1}D],$$

where

$$D_{ij} = \frac{(\alpha - \beta)a_i + (\alpha + \beta)a_j + 2a_i a_j}{(a_i + a_j)(\alpha + a_i)^2(\beta + a_i)(\alpha + a_j)}.$$

For $\alpha > \beta$ this matrix is positive definite. Thus, $\partial E(\alpha)/\partial\alpha$ is positive, and $E(\alpha)$ is strictly increasing. Thus, for $\alpha > \beta$, the $\bar{k}_i(\alpha)$ estimates are more efficient than the least square estimates.

5. Acknowledements. I am indebted to Professor H. B. Mann for suggesting the problem and to Professor Lucien LeCam for many valuable suggestions.

REFERENCES

[1] H. B. Mann, "A theory of estimation for the fundamental random process and the Ornstein Uhlenbeck process," *Sankhyā*, Vol. 13 (1954), p. 325.

[2] H. B. Mann and P. B. Moranda, "On the efficiency of the least square estimates of parameters in the Ornstein Uhlenbeck process," *Sankhyā*, Vol. 13 (1954), p. 351.

[3] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.

UNBIASED ESTIMATION OF CERTAIN CORRELATION COEFFICIENTS[1]

By Ingram Olkin[2] and John W. Pratt[3]

University of Chicago

1. Summary and introduction. This paper deals with the unbiased estimation of the correlation of two variates having a bivariate normal distribution (Sec. 2), and of the intraclass correlation, i.e., the common correlation coefficient of a p-variate normal distribution with equal variances and equal covariances (Sec. 3).

In both cases, the estimator has the following properties. It is a function of a complete sufficient statistic and is therefore the unique (except for sets of probability zero) minimum variance unbiased estimator. Its range is the region of possible values of the estimated quantity. It is a strictly increasing function of the usual estimator differing from it only by terms of order $1/n$ and consequently having the same asymptotic distribution.

Since the unbiased estimators are cumbersome in form in that they are expressed as series or integrals, tables are included giving the unbiased estimators as functions of the usual estimators.

In Sec. 4 we give an unbiased estimator of the squared multiple correlation. It has the properties mentioned in the second paragraph except that it may be negative, which the squared multiple correlation cannot.

In each case the estimator is obtained by inverting a Laplace transform.

We are grateful to W. H. Kruskal and L. J. Savage for very helpful comments and suggestions, and to R. R. Blough for his able computations.

2. Correlation coefficient. Let $(x_1, y_1), \cdots, (x_N, y_N)$ be independently distributed, each bivariate normal with means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation ρ. The problem is to estimate ρ unbiasedly in the cases (i) μ_1, μ_2 known, $\sigma_1^2, \sigma_2^2, \rho$ unknown, and (ii) all parameters unknown.

Sufficiency and invariance suggest that we confine ourselves to odd functions of r, where r is the usual sample correlation coefficient in either case, namely,

$$r = \frac{\sum (x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sqrt{\sum (x_i - \hat{\mu}_1)^2 \sum (y_i - \hat{\mu}_2)^2}},$$

where $(\hat{\mu}_1, \hat{\mu}_2)$ equals (μ_1, μ_2) in (i) and $(\bar{x}, \bar{y})$ in (ii).

2.1. *Derivation of the unbiased estimator.* The density of r is

$$p(r) = \frac{2^{n-2}}{\pi\Gamma(n-1)} (1-\rho^2)^{n/2}(1-r^2)^{(n-3)/2} \sum_{k=0}^{\infty} \Gamma^2\left(\frac{n+k}{2}\right)\frac{(2\rho r)^k}{k!}, \tag{2.1}$$

Received August 20, 1956; revised August 15, 1957.

[1] This research was sponsored by the Office of Ordnance Research, U. S. Army, and the Office of Naval Research.

[2] Work done while on leave from Michigan State University.

[3] Now at Harvard University.

where the degrees of freedom are $n = N$ and $N - 1$ in cases (i) and (ii). (We assume $n \geqq 2$, the case $n = 1$ being degenerate.) The condition $E[G(r)] = \rho$, i.e., $G(r)$ is unbiased, is equivalent to

$$\frac{2^{n-2}}{\pi\Gamma(n-1)} \sum_{k=0}^{\infty} \Gamma^2\left(\frac{n+k}{2}\right) \frac{(2\rho)^k}{k!} \int_{-1}^{1} G(r)(1-r^2)^{(n-3)/2} r^k \, dr = (1-\rho^2)^{-n/2}\rho$$

$$= \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{n}{2}+j\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{\rho^{2j+1}}{j!}.$$

Comparing coefficients of powers of ρ, we find that $G(r)$ is indeed an odd function, and that

$$\int_0^1 G(r)(1-r^2)^{(n-3)/2} r^{2j+1} \, dr = \frac{\pi\Gamma(n-1)\Gamma(2j+2)}{2^{n+2j}\Gamma^2\left(\frac{n+2j+1}{2}\right)} \frac{\Gamma\left(\frac{n+2j}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma(j+1)}.$$

Using the identity (e.g., [3, 12.4.4])

$$\sqrt{\pi}\,\Gamma(2p) = 2^{2p-1}\Gamma(p)\Gamma(p + 1/2),$$

and making the substitution $r = \exp(-\frac{1}{2}y)$, we obtain

$$\int_0^{\infty} G(e^{-\frac{1}{2}y})(1-e^{-y})^{(n-3)/2} e^{-y} e^{-jy} \, dy = \Gamma\left(\frac{n-1}{2}\right) \frac{\Gamma\left(\frac{3}{2}+j\right)\Gamma\left(\frac{n}{2}+j\right)}{\Gamma^2\left(\frac{n+1}{2}+j\right)}.$$

As a function of j, for $n \geqq 2$, the right-hand side is the unilateral Laplace transform of

$$e^{-\frac{1}{2}y}(1-e^{-y})^{((n-1)/2)-1} F\left(\frac{1}{2}, \frac{1}{2}; \frac{n-1}{2}; 1-e^{-y}\right)$$

[1, p. 262 (7)], where F is the hypergeometric function

$$F(\alpha, \beta; \gamma; x) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)\Gamma(\beta+k)\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma+k)} \frac{x^k}{k!}. \tag{2.2}$$

Therefore

$$G(r) = rF(\tfrac{1}{2}, \tfrac{1}{2}; (n-1)/2; 1-r^2). \tag{2.3}$$

Some alternative representations of $G(r)$ are

$$G(r) = r \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \int_0^1 \frac{t^{-1/2}(1-t)^{((n-2)/2)-1}}{[1-t(1-r^2)]^{1/2}} \, dt \tag{2.4}$$

and

$$(2.5) \qquad G(r) = r \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \int_0^\infty \frac{t^{-1/2}(1+t)^{-(n-2)/2}}{(1+tr^2)^{1/2}}\, dt$$

[2, 2.12 (1) and (5)].

2.2 *Properties of the unbiased estimator.* $G(r)$ is an odd function of r by (2.3), and is strictly increasing since, in (2.5), $r(1 + tr^2)^{-1/2}$ is strictly increasing in r for each value of t, $0 < t < \infty$. For $\rho = \pm 1$, $G(r) = r = \pm 1$ with probability 1, and consequently, $-1 \leqq G(r) \leqq 1$, which is the range of ρ.

As remarked before, $G(r)$ is the unique minimum variance unbiased estimator of ρ.

To obtain the asymptotic distribution of $G(r)$, we note that, by (2.2),

$$F(\alpha, \beta; \gamma; x) = 1 + x\, O(1/\gamma)$$

as $\gamma \to \infty$ (uniformly in x for x in any bounded set), so that $G(r) = r + O_p(1/n)$. Therefore $\sqrt{n}[G(r) - \rho]$ has the same asymptotic distribution as $\sqrt{n}[r - \rho]$, which is $N(O, (1 - \rho^2)^2)$, [3, p. 366].

In order to facilitate the use of the unbiased estimator $G(r)$, Table 1 gives $G(r)$ and (for easier interpolation) $G(r)/r$ for $r = 0(.1)\,1$ and $n = 2(2)\,30$. The computation was carried out by means of the recursive relation

$$xF(\tfrac{1}{2}, \tfrac{1}{2}; \gamma + 1; x)$$
$$= \left[1 - \frac{1}{(2\gamma - 1)^2}\right] [(2x - 1)F(\tfrac{1}{2}, \tfrac{1}{2}; \gamma; x) + (1 - x)F(\tfrac{1}{2}, \tfrac{1}{2}; \gamma - 1; x)],$$

[2, 2.8 (30)], together with the initial conditions

$$F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}; x) = 1/\sqrt{1 - x},$$
$$F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}; x) = \text{arc sin}\, \sqrt{x}/\sqrt{x},$$

[2, 2.8 (4) and (13)].

Approximations for $G(r)$ can be obtained from the expansion (2.2), which gives

$$(2.6) \qquad \frac{G(r)}{r} = 1 + \frac{1 - r^2}{2(n-1)} + \frac{9(1 - r^2)^2}{8(n^2 - 1)} + O(n^{-3}).$$

(2.6) gives $G(r)/r$ within .01 for $n \geqq 14$ or .001 for $n \geqq 36$ if two terms are included, and within .01 for $n \geqq 10$ or .001 for $n \geqq 18$ if three terms are included. The neglected terms in the first line of (2.6) are all positive and decreasing in r^2 and n. Therefore, if $G(r)$ is estimated by cutting off this series, the estimate will be too small, by a percentage which decreases as r^2 and n increase.

The k that minimizes the maximum over r of the absolute difference between (2.6) and $1 + (1 - r^2)/2(n - k)$ is, for large n, $(-7 + 9\sqrt{2})/2 = 2.87$. This suggests the approximation

$$(2.7) \qquad \frac{G(r)}{r} = 1 + \frac{1 - r^2}{2(n-3)}.$$

This is accurate within .01 for $n \geqq 8$, and within .001 for $n \geqq 18$.

TABLE 1

Ordinary bivariate correlation coefficient, n degrees of freedom

$$G(r) = rF\left(1/2, 1/2; \frac{n-1}{2}; 1 - r^2\right), \qquad r = \frac{s_{12}}{\sqrt{s_{11}\, s_{22}}}$$

1a. *Table of $G(r)$*

n	r										
	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	0	1	1	1	1	1	1	1	1	1	1
4	0	.148	.280	.398	.506	.605	.695	.780	.858	.931	1
6	0	.117	.232	.343	.450	.552	.650	.744	.833	.918	1
8	0	.110	.220	.327	.432	.534	.633	.730	.823	.913	1
10	0	.107	.214	.319	.423	.525	.625	.722	.817	.910	1
12	0	.106	.211	.315	.418	.520	.620	.718	.814	.908	1
14	0	.105	.209	.312	.415	.516	.616	.715	.812	.907	1
16	0	.104	.207	.311	.413	.514	.614	.713	.810	.906	1
18	0	.103	.206	.309	.411	.512	.612	.711	.809	.905	1
20	0	.103	.206	.308	.410	.511	.611	.710	.808	.905	1
22	0	.103	.205	.307	.409	.510	.610	.709	.807	.904	1
24	0	.102	.205	.307	.408	.509	.609	.708	.806	.904	1
26	0	.102	.204	.306	.407	.508	.608	.707	.806	.903	1
28	0	.102	.204	.305	.407	.507	.607	.707	.805	.903	1
30	0	.102	.204	.305	.406	.507	.607	.706	.805	.903	1
∞	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

1b. *Table of $G(r)/r$*

2	∞	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1
4	1.571	1.478	1.398	1.327	1.265	1.209	1.159	1.114	1.073	1.035	1
6	1.178	1.173	1.161	1.144	1.125	1.105	1.083	1.062	1.041	1.020	1
8	1.104	1.103	1.098	1.090	1.080	1.068	1.056	1.042	1.028	1.014	1
10	1.074	1.073	1.070	1.065	1.058	1.050	1.042	1.032	1.022	1.011	1
12	1.057	1.056	1.054	1.050	1.046	1.040	1.033	1.026	1.018	1.009	1
14	1.046	1.046	1.044	1.041	1.038	1.033	1.027	1.021	1.015	1.008	1
16	1.039	1.039	1.037	1.035	1.032	1.028	1.023	1.018	1.013	1.006	1
18	1.034	1.033	1.032	1.030	1.028	1.024	1.020	1.016	1.011	1.006	1
20	1.030	1.029	1.028	1.027	1.024	1.022	1.018	1.014	1.010	1.005	1
22	1.027	1.026	1.025	1.024	1.022	1.019	1.016	1.013	1.009	1.005	1
24	1.024	1.024	1.023	1.022	1.020	1.018	1.015	1.012	1.008	1.004	1
26	1.022	1.022	1.021	1.020	1.018	1.016	1.014	1.011	1.007	1.004	1
28	1.020	1.020	1.019	1.018	1.017	1.015	1.012	1.010	1.007	1.004	1
30	1.019	1.018	1.018	1.017	1.015	1.014	1.012	1.009	1.006	1.003	1
∞	1	1	1	1	1	1	1	1	1	1	1

By (2.2) and (2.3), $G(r)/r$ is larger than 1 and decreasing in r^2 and n, as Table 1b suggests.

2.3 *Partial correlation coefficient.* We observe that an unbiased estimator of the partial correlation coefficient can be immediately obtained from the preceding section. More precisely, suppose the columns of $X:p \times N$ are independently distributed each as p-variate normal with mean vector μ and covariance matrix Σ. We wish to give an unbiased estimator of the partial correlation coefficient $\rho_{12\cdot(q\cdots p)}$. The usual estimator, $r_{12\cdot(q\cdots p)}$, has the density (2.1) with $n = N - (p - q)$ if μ is known, and $n = N - 1 - (p - q)$ if μ is unknown. Therefore $G(r_{12\cdot(q\cdots p)})$ (with appropriate n) is the unique minimum variance unbiased estimator of $\rho_{12\cdot(q\cdots p)}$ and possesses the other properties of $G(r)$.

3. Intraclass correlation coefficient. Let the columns of X: $p \times N$ be independently distributed, each as $N(\mu, \Sigma^*)$, i.e., p-variate normal with mean vector μ and covariance matrix Σ^*. Suppose Σ^* is of the form $\sigma^2[(1 - \rho) I + \rho ee']$, where $e' = (1, \cdots, 1)$, i.e., $\sigma_{ii}^* = \sigma^2$, $\sigma_{ij}^* = \rho\sigma^2$ $(i \neq j)$, with ρ and σ^2 unknown. The problem is to estimate ρ unbiasedly.

We note that ρ is just the slope of the regression line of x_2 on x_1, and is therefore estimated unbiasedly by

$$\hat{\rho} = \frac{\sum_{\alpha=1}^{N} (x_{1\alpha} - x_{1\cdot})(x_{2\alpha} - x_{2\cdot})}{\sum_{\alpha=1}^{N} (x_{1\alpha} - x_{1\cdot})^2},$$

where a dot indicates an average over the omitted subscript. We will see presently that there is a complete sufficient statistic (u, v). $\hat{\rho}$ is not a function of (u, v), nor is it confined to the range of ρ, namely, $-1/(p - 1)$ to 1. However, by the Blackwell-Rao theorem, $E(\hat{\rho} \mid u, v)$ is the unique minimum variance unbiased estimator of ρ. Since $E(\hat{\rho} \mid u, v)$ is difficult to obtain, we shall use the joint distribution of u and v to obtain an unbiased estimator $h(u, v)$ of ρ, which, by completeness, must equal $E(\hat{\rho} \mid u, v)$.

As in the previous section, sufficiency and invariance suggest that we confine ourselves to functions of the conventional estimator r' of ρ. However, it is easier to deal with the density of (u, v), and it will turn out that the unbiased estimator $h(u, v)$ is a function $H(r')$ of r' alone.

3.1 *Reduction to canonical form.* Let $\Delta:p \times p$ be an orthogonal matrix with first row $p^{-1/2}e'$, and let $Y = \Delta X$. Then the columns of Y are independently distributed, each as $N(\Delta\mu, \Delta\Sigma^*\Delta')$. Now

$$\Delta\Sigma^*\Delta' = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 I \end{pmatrix},$$

where $\sigma_1^2 = \sigma^2 [1 + (p - 1)\rho]$, $\sigma_2^2 = \sigma^2 (1 - \rho)$. Because of the particular diagonal form of the covariance matrix, the $y_{i\alpha}(i = 1, \cdots, p; \alpha = 1, \cdots, N)$ are independent, and if we let $\eta = \Delta\mu = Ey$, then $y_{1\alpha}$ is $N(\eta_1, \sigma_1^2)$, $(\alpha = 1, \cdots, N)$ and $y_{i\alpha}$ is $N(\eta_i, \sigma_2^2)$, $(i = 2, \cdots, p; \alpha = 1, \cdots, N)$. We can therefore obtain two sums of squares, u and v, sufficient for σ_1^2 and σ_2^2 and distributed independently as $\sigma_1^2 \chi_a^2$ and $\sigma_2^2 \chi_b^2$ where the degrees of freedom a and b depend on our knowledge of μ. To write u and v conveniently, we first observe that

$$\sum_{i=1}^{p} \sum_{\alpha=1}^{N} y_{i\alpha}^2 = \operatorname{tr} YY' = \operatorname{tr} XX' = \sum_{i=1}^{p} \sum_{\alpha=1}^{N} x_{i\alpha}^2,$$

$$\sum_{\alpha=1}^{N} y_{1\alpha}^2 = p^{-1}e'XX'e = p \sum_{\alpha=1}^{N} x_{\cdot\alpha}^2,$$

$$\sum_{i=1}^{p} y_{i\cdot}^2 = N^{-2}e'Y'Ye = N^{-2}e'X'Xe = \sum_{i=1}^{p} x_{i\cdot}^2,$$

$$y_{1\cdot} = N^{-1}p^{-1/2}e'Xe = p^{1/2}x_{\cdot\cdot}.$$

Precisely, we consider the following three cases:

(i) $\mu = 0$ and hence $\eta = \Delta\mu = 0$. Let $a = N$, $b = (p-1)N$,

$$u = \sum_{\alpha=1}^{N} y_{1\alpha}^2 = p \sum_{\alpha=1}^{N} x_{\cdot\alpha}^2,$$

$$v = \sum_{i=2}^{p} \sum_{\alpha=1}^{N} y_{i\alpha}^2 = \sum_{\alpha=1}^{N} \sum_{i=1}^{p} (x_{i\alpha} - x_{\cdot\alpha})^2.$$

(ii) μ completely unknown and hence $\eta = \Delta\mu$ is also completely unknown. Let $a = N - 1$, $b = (p-1)(N-1)$,

$$u = \sum_{\alpha=1}^{N} (y_{1\alpha} - y_{1\cdot})^2 = p \sum_{\alpha=1}^{N} (x_{\cdot\alpha} - x_{\cdot\cdot})^2,$$

$$v = \sum_{i=2}^{p} \sum_{\alpha=1}^{N} (y_{i\alpha} - y_{i\cdot})^2 = \sum_{\alpha=1}^{N} \sum_{i=1}^{p} (x_{i\alpha} - x_{i\cdot} - x_{\cdot\alpha} + x_{\cdot\cdot})^2.$$

(iii) $\mu = \omega e$, where ω is an unknown scalar, and hence $\eta = \omega\Delta e = \omega\sqrt{p}(1, 0, \cdots, 0)'$. Let $a = N - 1$, $b = (p-1)N$,

$$u = \sum_{\alpha=1}^{N} (y_{1\alpha} - y_{1\cdot})^2 = p \sum_{\alpha=1}^{N} (x_{i\alpha} - x_{\cdot\cdot})^2,$$

$$v = \sum_{i=2}^{p} \sum_{\alpha=1}^{N} y_{i\alpha}^2 = \sum_{i=1}^{p} \sum_{\alpha=1}^{N} (x_{i\alpha} - x_{\cdot\alpha})^2.$$

In each case u/σ_1^2, and v/σ_2^2 are independently distributed as χ_a^2 and χ_b^2, and it is easily shown that (u, v) is a complete sufficient statistic for (σ_1^2, σ_2^2). The three cases can thus be treated simultaneously.

3.2 *Derivation of the unbiased estimator.* The condition that $h(u, v)$ be unbiased is

$$\int_0^\infty \int_0^\infty h(u, v) u^{a/2-1} v^{b/2-1} e^{-\theta u - \phi v} \, du \, dv \tag{3.1}$$
$$= \Gamma\left(\frac{a}{2}\right) \Gamma\left(\frac{b}{2}\right) \frac{\phi - \theta}{\phi + (p-1)\theta} \cdot \frac{1}{\theta^{a/2}\phi^{b/2}},$$

where $\theta = 1/(2\sigma_1^2)$, $\phi = 1/(2\sigma_2^2)$. The right-hand side is the bivariate Laplace transform of

$$
(3.2)\qquad \begin{aligned} &\left(\frac{b}{2} - 1\right)\int_0^L [u - (p-1)y]^{\frac{a}{2}-1}(v-y)^{\frac{b}{2}-2}\,dy \\ &\qquad - \left(\frac{a}{2} - 1\right)\int_0^L [u - (p-1)t]^{\frac{a}{2}-1}(v-t)^{\frac{b}{2}-1}\,dt, \end{aligned}
$$

where $L = \min\,[u/(p-1), v]$, [4, p. 36 (Satz 12), p. 208 (9), p. 236 (87)]. Integrating the first term of (3.2) by parts and letting $z = u/[(p-1)v]$, we obtain

$$
(3.3)\qquad \begin{aligned} h(u, v) = h^*(z) &= 1 - \left(\frac{a}{2} - 1\right)\frac{p}{p-1}\int_0^1 (1 - zw)^{\frac{b}{2}-1}(1-w)^{\frac{a}{2}-2}\,dw \\ &= 1 - \frac{p}{p-1}F\left(1, 1 - \frac{b}{2}; \frac{a}{2}; z\right) \qquad \text{for} \quad 0 \leqq z \leqq 1, \end{aligned}
$$

$$
(3.4)\qquad \begin{aligned} h(u, v) = h^*(z) &= 1 - \left(\frac{a}{2} - 1\right)\frac{p}{p-1}\frac{1}{z}\int_0^1 \left(1 - \frac{1}{z}w\right)^{\frac{a}{2}-2}(1-w)^{\frac{b}{2}-1}\,dw \\ &= 1 - \frac{2}{b}\left(\frac{a}{2} - 1\right)\frac{p}{p-1}\frac{1}{z}F\left(1, 2 - \frac{a}{2}; \frac{b}{2} + 1; \frac{1}{z}\right) \\ &\qquad\qquad \text{for} \quad z \geqq 1, \end{aligned}
$$

[2, 2.12 (1)]. Integrating the second term of (3.2) by parts we obtain the following alternative to (3.4):

$$
(3.5)\qquad \begin{aligned} h^*(z) &= \left(\frac{b}{2} - 1\right)\frac{p}{p-1}\int_0^1 \left(1 - \frac{1}{z}w\right)^{\frac{a}{2}-1}(1-w)^{\frac{b}{2}-2}\,dw - \frac{1}{p-1} \\ &= \frac{p}{p-1}F\left(1, 1 - \frac{a}{2}; \frac{b}{2}; \frac{1}{z}\right) - \frac{1}{p-1} \qquad \text{for} \quad z \geqq 1, \end{aligned}
$$

[2, 2.12 (1)].

The conventional estimate of ρ is (e.g., [5] and [6]),

$$
(3.6)\qquad \begin{aligned} r' &= \frac{p}{p-1}\,\frac{\sum_\alpha \sum_{i \neq j} (x_{i\alpha} - \hat{\mu}_i)(x_{j\alpha} - \hat{\mu}_j)}{\sum_\alpha \sum_i (x_{i\alpha} - \hat{\mu}_i)^2} \\ &= \frac{1}{p-1}\left(\frac{pu}{u+v} - 1\right) = \frac{pz}{1 + (p-1)z} - \frac{1}{p-1}, \end{aligned}
$$

where $\hat{\mu}$ is the appropriate estimate of μ in (i), (ii), or (iii). Now

$$
(3.7)\qquad z = \frac{(p-1)\,r' + 1}{(p-1)^2(1 - r')},
$$

which is a strictly increasing function of r'. Thus $h^*(z)$ is a function $H(r')$ of r'.

TABLE 2

Intraclass correlation coefficient, bivariate case, n degrees of freedom

$$H(r') = r'F(1/2, 1; n/2; 1 - r'^2)$$

2a. *Table of* $H(r')$

n	r'										
	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	0	1	1	1	1	1	1	1	1	1	1
4	0	.182	.333	.462	.571	.667	.750	.824	.889	.947	1
6	0	.132	.259	.379	.490	.593	.688	.775	.856	.931	1
8	0	.120	.237	.351	.459	.563	.661	.753	.841	.923	1
10	0	.114	.227	.337	.444	.547	.647	.741	.832	.918	1
12	0	.111	.221	.329	.435	.538	.638	.734	.826	.915	1
14	0	.109	.217	.324	.429	.532	.631	.728	.822	.913	1
16	0	.108	.215	.321	.425	.527	.627	.724	.819	.911	1
18	0	.107	.213	.318	.422	.524	.624	.722	.817	.910	1
20	0	.106	.211	.316	.419	.521	.621	.719	.815	.909	1
22	0	.105	.210	.314	.417	.519	.619	.717	.814	.908	1
24	0	.105	.209	.313	.416	.517	.617	.716	.813	.907	1
26	0	.104	.208	.312	.414	.516	.616	.715	.812	.907	1
28	0	.104	.208	.311	.413	.515	.615	.713	.811	.906	1
30	0	.104	.207	.310	.412	.513	.614	.713	.810	.906	1
∞	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

2b. *Table of* $H(r')/r'$

2	∞	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1
4	2.000	1.818	1.667	1.538	1.429	1.333	1.250	1.176	1.111	1.053	1
6	1.333	1.322	1.296	1.262	1.224	1.185	1.146	1.107	1.070	1.034	1
8	1.200	1.196	1.185	1.169	1.149	1.126	1.102	1.076	1.051	1.025	1
10	1.143	1.141	1.134	1.124	1.110	1.095	1.078	1.059	1.040	1.020	1
12	1.111	1.110	1.105	1.098	1.088	1.076	1.062	1.048	1.033	1.017	1
14	1.091	1.090	1.086	1.080	1.073	1.063	1.052	1.041	1.028	1.014	1
16	1.077	1.076	1.073	1.068	1.062	1.054	1.045	1.035	1.024	1.012	1
18	1.067	1.066	1.063	1.059	1.054	1.047	1.040	1.031	1.021	1.011	1
20	1.059	1.058	1.056	1.053	1.048	1.042	1.035	1.027	1.019	1.010	1
22	1.053	1.052	1.050	1.047	1.043	1.038	1.032	1.025	1.017	1.009	1
24	1.048	1.047	1.045	1.043	1.039	1.034	1.029	1.023	1.016	1.008	1
26	1.043	1.043	1.042	1.039	1.036	1.032	1.027	1.021	1.014	1.007	1
28	1.040	1.040	1.038	1.036	1.033	1.029	1.024	1.019	1.013	1.007	1
30	1.037	1.037	1.035	1.033	1.031	1.027	1.023	1.018	1.012	1.006	1
∞	1	1	1	1	1	1	1	1	1	1	1

3.3. *Properties of the unbiased estimator.* For $\rho = 1$, $z = \infty$, $r' = 1$ with probability 1, and $h^*(\infty) = H(1) = 1$; for $\rho = -1/(p-1)$, $z = 0$, $r' = -1/(p-1)$ with probability 1, and $h^*(0) = H(-1/(p-1)) = -1/(p-1)$. Thus in the two cases when Σ^* is singular, $h^*(z) = H(r') = \rho$ with probability 1. Furthermore, $h^*(z)$ is a strictly increasing function of z, since the integrand of (3.3) for $0 \leqq z \leqq 1$ and of (3.5) for $z \geqq 1$ is strictly monotone for each value of w, $0 < w < 1$. Consequently, $H(r')$ is a strictly increasing function of r' and $-1/(p-1) \leqq h^*(z) = H(r') \leqq 1$, which is the range of ρ.

As remarked before, $h^*(z) = H(r')$ is the unique minimum variance unbiased estimator of ρ.

We will now obtain the asymptotic distribution of $h^*(z)$. Note that z is distributed as

$$\frac{1+(p-1)\rho}{1-\rho}\,\frac{a}{b(p-1)}\,F_{a,b},$$

and that $\sqrt{(p-1)N/p}\,(F_{a,b}-1)$ is asymptotically $N(0, 1)$. Therefore, letting $z_0 = [1+(p-1)\rho]/[(p-1)^2(1-\rho)]$, the quantity,

$$\sqrt{\frac{(p-1)N}{p}}\left[z\,\frac{(p-1)^2(1-\rho)}{1+(p-1)\rho}-1\right] = \sqrt{\frac{N}{p}}\,\frac{(p-1)^{5/2}(1-\rho)}{1+(p-1)\rho}\,(z-z_0)$$

is asymptotically $N(0, 1)$. But, by (3.3), denoting $N^{-5/6}$ by ϵ, we have, for $z \leqq 1$,

$$\begin{aligned}1-h^*(z) &= \frac{p}{p-1}\,\frac{N}{2}\int_0^{\epsilon}[1-w-(p-1)zw]^{N/2}\,dw\\ &\qquad\cdot[1+O(\epsilon^2)]^{N/2}[1+O(\epsilon)]+NO(1-\epsilon)^{N/2}\\ &= \frac{p}{p-1}\,\frac{1}{1+(p-1)z}+O\left(\frac{1}{N}\right),\end{aligned}$$

uniformly in z. We obtain the same result for $z \geqq 1$ from (3.4). Therefore

$$h^*(z) = \rho+(p-1)^2(1-\rho)^2(z-z_0)/p+O_p(1/N).$$

Therefore

$$\sqrt{N}\,[h^*(z)-\rho] = \sqrt{N}\,[H(r')-\rho] \quad \text{is asymptotically} \quad N(0, \sigma^2),$$

where $\sigma^2 = (1-\rho)^2[1+(p-1)\rho]^2/[p(p-1)]$.

Expanding r' about z_0 in (3.6) we find

$$r' = h^*(z)+O_p(1/N) = H(r')+O_p(1/N),$$

so that r' is asymptotically equivalent to $H(r')$. Incidentally, we find that $\sqrt{N}(r'-\rho)$ is asymptotically $N(0, \sigma^2)$, with the same σ^2.

In order to facilitate the use of the unbiased estimator in the bivariate case with n degrees of freedom, i.e., case (i) or (ii) with $p = 2$, Table 2 gives $H(r')$ and, (for easier interpolation), $H(r')/r'$ for $r' = 0(.1)1$ and $a = b = n = 2(2)\,30$. In this case, $H(r') = H_n(r')$ is an odd function of r'. The computation was carried out by means of the recursive relation

$$H_n(r') = \frac{n-2}{n-3}\left[\frac{r'}{1-r'^2}-\frac{r'^2}{1-r'^2}\,H_{n-2}(r')\right],$$

together with the initial conditions

$$H_2(r') = 1, \qquad H_4(r') = \frac{2r'}{1+r'}, \qquad\qquad (r' > 0).$$

$H_n(r')$ was derived for $n \geqq 3$. For $n = 2$, the inversion in (3.1) must be carried out separately, and the result agrees with the final form of $h^*(z)$ in (3.4) and (3.6). The recursive relation is obtained by application of the relations [2, 2.8 (36) and (39)]. The same formulas give recursive relations for any values of p, a, b.

For the bivariate case with n degrees of freedom,

$$H(r') = r'F(\tfrac{1}{2}, 1; n/2; 1 - r'^2), \tag{3.8}$$

which is obtained from [2, 2.8 (36) and 2.11 (34)].

Approximations for $H(r')$ can be obtained from the expansion (2.2) applied to (3.8), which gives

$$\frac{H(r')}{r'} = 1 + \frac{1 - r'^2}{n} + \frac{3(1 - r'^2)^2}{n(n+2)} + O(n^{-3}). \tag{3.9}$$

This gives $H(r')/r'$ within .01 for $n \geqq 19$ or .001 for $n \geqq 57$ if two terms are included, and within .01 for $n \geqq 12$ or .001 for $n \geqq 26$ if three terms are included. As in (2.6), the neglected terms in (3.9) are all positive and decreasing in r'^2 and n.

The k that minimizes the maximum over r of the absolute difference between $H(r')/r'$ and $1 + (1 - r'^2)/(n - k)$ is, for large n, $6(-1 + \sqrt{2}) = 2.48$. This suggests the approximation

$$\frac{H(r')}{r'} = 1 + \frac{1 - r'^2}{n - 5/2}. \tag{3.10}$$

This is accurate within .01 for $n \geqq 10$ or .001 for $n \geqq 26$.

4. Multiple correlation coefficient. Suppose we have N independent observations on a $p + 1$-variate normal distribution with mean vector μ and covariance matrix Σ, and we wish to give an unbiased estimator of the squared multiple correlation

$$\rho^2 = \rho^2_{0.(12\cdots p)} = 1 - \mathcal{R}/\mathcal{R}_{00},$$

where $\mathcal{R}$ is the determinant of the correlation matrix and $\mathcal{R}_{00}$ is its first cofactor. We are concerned with the cases (i) μ known, Σ unknown, and (ii) all parameters unknown.

As in 2.1, we confine ourselves to functions of

$$r^2 = r^2_{0.(12\cdots p)} = 1 - R/R_{00},$$

where R is the determinant of the appropriate (to (i) or (ii)) sample correlation matrix and R_{00} is its first cofactor.

The condition that $I(r^2)$ be unbiased is

$$\sum_{k=0}^{\infty} \frac{\Gamma^2\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{p}{2} + k\right)} \frac{\rho^{2k}}{k!} \int_0^1 I(r^2)(r^2)^{((p-2)/2)+k}(1 - r^2)^{(n-p-1)/2}\, dr^2$$
$$= \Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{n}{2}\right)(1 - \rho^2)^{-n/2}\rho^2,$$

where $n = N$ and $(N - 1)$ in cases (i) and (ii). Following the methods of Sec. 2, we obtain

$$I(r^2) = 1 - \frac{n-2}{n-p}(1 - r^2)F\left(1, 1; \frac{n-p+2}{2}; 1 - r^2\right).$$

As usual, $I(r^2)$ is strictly increasing in r^2, and differs from it only by terms of order $1/N$, and it is the unique minimum variance unbiased estimator of ρ^2. Also $I(1) = 1$. However, $I(0) = -p/(n - p - 2)$. We cannot hope for a non-negative unbiased estimator, since there is no region in the sample space having zero probability for $\rho^2 = 0$ and positive probability for $\rho^2 > 0$. For the same reason there can be no positive unbiased estimator of ρ either.

REFERENCES

[1] Bateman Manuscript Project, California Institute of Technology, *Tables of Integral Transforms*, Vol. 1, A. Erdelyi, editor, McGraw-Hill Book Co., New York, 1954.

[2] Bateman Manuscript Project, California Institute of Technology, *Higher Transcendental Functions*, Vol. 1, A. Erdelyi, editor, McGraw-Hill Book Co., New York, 1953.

[3] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[4] G. Doetsch and Dietrich Voelker, *Die Zweidimensionale Laplace-Transformation*, Birkhauser Basel, Zurich, 1950.

[5] D. B. DeLury, "Note on correlations," *Ann. Math. Stat.*, Vol. 9 (1933), pp. 149–151.

[6] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin, London, 1947.

TRANSIENT ATOMIC MARKOV CHAINS WITH A DENUMERABLE NUMBER OF STATES[1]

By Leo Breiman

University of California

1. Introduction. Many of the more interesting transient Markov chains have the property that for any set of states A and any initial distribution, the probability of entering A infinitely often (i.o.) is either zero always or one always. This type of chain has been termed *atomic* by D. Blackwell [1] and is exemplified by the three-dimensional random walk or by the successive sums of independent, identically distributed random variables.

In this paper we investigate the "fine structure" of an atomic chain, that is, we try to characterize the class of all sets A such that $P(x_n \varepsilon A \text{ i.o.}) = 0$. The study is restricted to atomic chains with a countable set of states which, for convenience of notation, we identify with the integers, and with stationary transition probabilities $p_{ij}^{(n)}$.

The martingale convergence theorem is used in [1] to show that a necessary and sufficient condition for atomicity is that every bounded solution ϕ of

$$\phi(i) = \sum_j p_{ij}\phi(j)$$

be constant. We use as our main tool the semi-martingale convergence theorem and the corresponding equation $\phi(i) \geqq \sum_j p_{ij}\phi(j)$ and obtain a complete, but not simple, characterization of the fine structure of transient atomic chains.

To illustrate the use of the above characterization we prove two theorems regarding the return to equilibrium times $x_0, x_1, \cdots$ in the coin-tossing game. The latter of these is then used to prove that there exists no set of numbers $\{\lambda_m\}$ such that[2] $P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m \varepsilon A} \lambda_m < \infty$.

This last result shows that, in general, there is no simple resolution to the question of defining the fine structure. There are, however, a number of interesting transient atomic chains which have the property that every infinite set of states is entered infinitely often with probability one. These chains are the subject of papers by Chung and Derman [2], and Breiman [3].

2. Use of the semi-martingale theorem.

Theorem 1. *Let $x_0, x_1, \cdots$ be an atomic chain. Then for ϕ any nonnegative solution of*

$$\text{(a)} \qquad \phi(i) \geqq \sum_j p_{ij}\phi(j)$$

---

[1] This paper was prepared with the support of the Office of Ordnance Research, U.S. Army, under Contract DA-04-200-ORD-171.

[2] The referee has informed us that a similar theorem for the three-dimensional random walk has been proved by P. Erdös and B. J. Murdoch (unpublished).

which is finite for at least one value of i, $\phi(x_n)$ *converges almost surely (a.s.) to a constant independent of the initial distribution.*

PROOF. Let ϕ be a nonnegative solution of (a) with $\phi(0) < \infty$, and let R be the set of all states i such that P (entering $i \mid x_0 = 0) > 0$. From the atomicity and $P(x_n \varepsilon \tilde{R} \text{ i.o.} \mid x_0) = 0$, where $\tilde{R}$ is the complement of R, follows $P(x_n \varepsilon \tilde{R}$ i.o.$) = 0$. Therefore, it is sufficient to prove the theorem for initial distributions concentrated on R noting that ϕ is finite on R. Pick any such distribution $\{p_j\}$ such that $\sum_j p_j\phi(j) < \infty$ and $p_j > 0$, all $j \varepsilon R$. The random variables $\phi(x_n)$ form a semi-martingale with respect to the fields generated by $x_0, x_1, \cdots$, since

$$E(\phi(x_n) \mid x_{n-1}, \cdots, x_0) = E(\phi(x_n) \mid x_{n-1}) = \sum p_{x_{n-1}j}\phi(j) \leqq \phi(x_{n-1}),$$

$$E|\phi(x_n)| = E\phi(x_n) \leqq E\phi(x_0) < \infty.$$

By the semi-martingale theorem [4], $\phi(x_n)$ converges a.s. Suppose this limit is nonconstant, then there will be a number $a > 0$ such that if A is the set of states defined by $\{j; \phi(j) \geqq a\}$, then $0 < P(x_n \varepsilon A \text{ i.o.}) < 1$. Hence the limit is constant, and since $\phi(x_n)$ must converge to this same constant with the initial distribution concentrated at any single state in R, the theorem is valid.

We note that the same result is true for ϕ any bounded solution of (a) because for any sufficiently large constant α, $\phi + \alpha$ is a positive solution.

A simple but informative corollary of the above theorem demonstrates the special applicability of (a) to the transient case.

COROLLARY. *All the states of an atomic chain are recurrent if and only if all bounded solutions of* (a) *are constant. For a transient atomic chain there is at least one nonconstant bounded solution of* (a)

PROOF. Let $\phi(i) = P$ (entering $i_0 \mid x_0 = i$), so that $\phi(i_0) = 1$, and

$$\phi(i) = E(P(\text{entering } i_0 \mid x_1, x_0) \mid x_0 = i) \geqq E(\phi(x_1) \mid x_0 = i) = \textstyle\sum_j p_{ij}\phi(j).$$

If every solution of (a) is constant, then for every i_0 we have $P(\text{entering } i_0 \mid x_0 = i) = 1$. This implies that return to every state is certain. Now assume that every state is recurrent and let ϕ be any bounded solution of (a). If $\phi(i) \neq \phi(j)$, $\phi(x_n)$ cannot possibly converge to a constant since both i and j are entered i.o. with probability one. If there are transient states present the function ϕ defined above cannot be constant for all i_0.

3. Characterization of the fine structure. We use the notation

$$u_{ik} = E(\text{number of visits to } k \mid x_0 = i),$$

$$u_{ik}^{(n)} = \delta_{ik} + p_{ik} + \cdots + p_{ik}^{(n-1)} \qquad \delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

and recall that $u_{ik} = \lim_n u_{ik}^{(n)}$.

THEOREM 2. *If* $x_0, x_1, \cdots$ *is an atomic chain, then for every nonnegative sequence of numbers* $\{\alpha_k\}$ *with* $\sum_k u_{0k}\alpha_k < \infty$ *and every* $\epsilon > 0$, *the set of states* $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq \epsilon\}$ *has the property* $P(x_n \epsilon A_\alpha \text{ i.o.}) = 0$. *Conversely, every* *set of states A such that $P(x_n \varepsilon A$ i.o.$) = 0$ is included in at least one of the sets A_α as defined above.*

PROOF. Let $\{\alpha_k\}$ be a sequence fulfilling the conditions of the theorem and let $\phi(i) = \sum_k u_{ik}\alpha_k$. The identity $\sum_j p_{ij}u_{jk} = u_{ik} - \delta_{ik}$ leads to the equation $\Sigma_j p_{ij}\phi(j) = \phi(i) - \alpha_i$. Thus, theorem 1 applies and $\phi(x_n)$ converges a.s. to a constant. Since the properties in which we are interested do not, in an atomic chain, depend on initial conditions, it is sufficient to take $x_0 = 0$. Then, iterating the equation which ϕ satisfies,

$$E(\phi(x_n) \mid x_0 = 0) = \sum_k (u_{0k} - u_{0k}^{(n)})\alpha_k \to 0$$

and by a semi-martingale inequality ([4], p. 325) which states that

$$E(\text{a.s. limit}) \leqq E\phi(x_n)$$

we are able to conclude that the a.s. limit of $\phi(x_n)$ is identically zero. This implies that $P(\phi(x_n) \geqq \epsilon$ i.o.$) = 0$ and proves one part of the theorem.

To get the second part, let A be any set of states with $P(x_n \varepsilon A$ i.o.$) = 0$. Form the function $\phi(i) = P(\text{entering } A \mid x_0 = i)$, so that $\phi(i) = 1$, all $i \varepsilon A$. It is easy to verify that ϕ satisfies (a), and thus $\phi(x_n)$ converges a.s. to some constant. We deduce that this constant is zero by noting that P(entering A after $n - 1$ steps) $= E\phi(x_n)$. Since $P(x_n \varepsilon A$ i.o.$) = 0$ we conclude that $E\phi(x_n) \to 0$ and apply the bounded convergence theorem to get the result. Let the nonnegative sequence $\{\alpha_i\}$ be defined by $\phi(i) = \alpha_i + \sum_j p_{ij}\phi(j)$. Iterating this equation

$$\phi(i) = \sum_j p_{ij}^{(n)}\phi(j) + \sum_j u_{ij}^{(n)}\alpha_j.$$

By the boundedness of ϕ the second sum converges to $\sum_j u_{ij}\alpha_j$. The first sum must also converge to some bounded limit sequence $\{\lambda(i)\}$. Since

$$\lambda(i) = \sum_j p_{ij}\lambda(j),$$

by Blackwell's theorem as quoted above this sequence is constant, and by the convergence of $\phi(x_n)$ to zero, $\lambda(i) \equiv 0$. The set A is contained in the set $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq 1\}$ which proves the theorem.

4. Two theorems concerning the coin-tossing game. We apply theorem 2 to the Markov chain $x_0, x_1, \cdots$ whose values are the successive times of return to equilibrium in the fair coin-tossing game. The set of states is the set of all nonnegative even integers and we use the fact that this chain, being the sum of independent and identically distributed random variables, is atomic. It is well known that

$$\begin{aligned} u_{ik} &= 0, && k < i, \\ &\sim \frac{c}{\sqrt{(k - i)}}, && k \gg i. \end{aligned}$$

As it is evident that the characterization given in theorem 2 is invariant under asymptotic equivalence, we use $1/\sqrt{k - i}$ throughout this section in place of u_{ik} with the convention $\sqrt{0} = 1$ and $1/\sqrt{-} = 0$.

The first theorem we prove is similar to a theorem stated by Chung and Erdös [5].

THEOREM 3. *Let the sequence of even positive integers* $\{m_i\}$ *be such that the sequence* $\{\Delta_i\}$ *defined by* $\Delta_i = m_i - m_{i-1}$ *is nondecreasing. Then*

$$P(x_n \, \varepsilon \, \{m_i\} \text{ i.o.}) = 0 \Leftrightarrow \sum_i \frac{1}{\sqrt{m_i}} < \infty.$$

PROOF. If $\sum_i 1/\sqrt{m_i} < \infty$, the assertion follows immediately from the Borel-Cantelli lemma. Now assume that $P(x_n \, \varepsilon \, \{m_i\} \text{ i.o.}) = 0$, but that $\sum_i 1/\sqrt{m_i} = \infty$. By theorem 2, there is a nonnegative sequence $\{\alpha_i\}$ such that $\sum_k \alpha_k / \sqrt{k} < \infty$ and $\{m_i\} \subset \{i; \sum_k \alpha_k / \sqrt{k - i} \geqq \epsilon\}$. From this we have for all m_i

$$\sum_{k \geqq m_i} \frac{\alpha_k}{\sqrt{k - m_i}} \geqq \epsilon.$$

Define $\lambda_k^{(N)}$ by

$$\lambda_k^{(N)} = \frac{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}\,\sqrt{k - m_i}}}{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}}}.$$

It is evident that $\sum_k \lambda_k^{(N)} \alpha_k \geqq \epsilon$, all N, and that $\lim_N \lambda_k^{(N)} = 0$ for k fixed. We will show that $\lambda_k^{(N)} < c/\sqrt{k}$, all k, N, and conclude from the bounded convergence theorem the contradiction that $\lim_N \sum_k \lambda_k^{(N)} \alpha_k = 0$. To begin with, assume that $k \geqq m_N$, then

$$\sqrt{k}\,\lambda_k^{(N)} \leqq \frac{\displaystyle\sqrt{m_N} \sum_{i=0}^{N} \frac{1}{\sqrt{m_i}\,\sqrt{m_N - m_i}}}{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}}}.$$

By splitting the top sum into the two parts $m_i \leqq m_N/2$, $m_i > m_N/2$ and using our assumption concerning Δ_i, we get $\sqrt{k}\lambda_k^{(N)} \leqq 4$. Now if $k \leqq m_N$, let m_n be the largest of the m_i which is $\leqq k$. With this

$$\sqrt{k}\,\lambda_k^{(N)} \leqq \frac{\displaystyle\sqrt{m_n} \sum_{i=0}^{N} \frac{1}{\sqrt{m_i}\,\sqrt{m_n - m_i}}}{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}}}$$

and repeating the above argument results again in $\sqrt{k}\lambda_k^{(N)} \leqq 4$.

It is clear that in the above context, a little more attention to the appropriate inequalities would result in a considerable weakening of the growth condition on the sequence $\{m_i\}$.

We can get a result in another direction by combining our characterization with different inequalities. Let all the states between and including n_1 and n_2, $n_2 \geqq n_1$, be called an interval and denoted by $[n_1, n_2]$.

THEOREM 4. *If the sequence of disjoint finite intervals* $\{I_j\}$, $I_j = [m_j, M_j]$ *is such that for some* $\delta > 0$, $m_{j+1} \geqq (1+\delta)M_j$, *then, denoting*

$$l_j = M_j - m_j + 2,$$

$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0 \Leftrightarrow \sum_j \sqrt{l_j/M_j} < \infty.$$

PROOF. Define a sequence of intervals $I_j' = [m_j', M_j']$ by $m_j' = M_j$, $M_j' = M_j + \sqrt{l_j}$, where $\sqrt{l_j}$ is here to be interpreted as the greatest even integer less than $\sqrt{l_j}$. Let $\sum_j \sqrt{l_j/M_j} < \infty$ and define α_k by

$$\alpha_k = \begin{cases} 1 & \text{if } k \varepsilon \bigcup_j I_j', \\ 0 & \text{otherwise.} \end{cases}$$

By these definitions

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \leqq \frac{1}{2} \sum_j \sqrt{l_j/M_j} < \infty.$$

Thus the set $A = \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \frac{1}{2}\}$ has the property that $P(x_n \varepsilon A \ i.\ o.) = 0$. The set A includes, in particular, the integers i such that $i \leqq M_j$ and

$$\tfrac{1}{2} \leqq \sum_{k\varepsilon I_j'} \frac{1}{\sqrt{k-i}} \leqq \tfrac{1}{2}(\sqrt{M_j' - i} - \sqrt{m_j' - i}).$$

This inequality can be easily shown to be satisfied by all $i \geqq m_j$, which proves the theorem going one way.

To go the other way, assume that $P(x_n \epsilon \bigcup_j I_j \ i.\ o.) = 0$. Then there is a non-negative sequence α_k such that $\sum_k \alpha_k/\sqrt{k} < \infty$ and $\bigcup_j I_j \subset \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \epsilon\}$, from which, if $i \varepsilon I_j$, then $\sum_{k \geqq m_j} \alpha_k/\sqrt{k-i} \geqq \epsilon$. We wish to conclude that part of this sum is negligible and argue that if $i \varepsilon I_j$, and if j is sufficiently large

$$\sum_{k \geqq m_{j+1}} \sqrt{\frac{k}{k-i}} \frac{\alpha_k}{\sqrt{k}} \leqq \sqrt{\frac{m_{j+1}}{m_{j+1} - M_j}} \sum_{k \geqq m_{j+1}} \frac{\alpha_k}{\sqrt{k}} \leqq \frac{\epsilon}{2}$$

so that if $i \varepsilon I_j$,

$$\sum_{m_{j+1} \geqq k \geqq m_j} \frac{\alpha_k}{\sqrt{k-i}} \geqq \frac{\epsilon}{2}.$$

We sum this last inequality over $i \varepsilon I_j$ to get

$$\sum_{m_{j+1} > k \geqq m_j} \alpha_k \left(\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k-i}} \right) \geqq \frac{\epsilon}{4} l_j.$$

It can be easily shown that

$$\sum_{i\varepsilon I_j} \frac{1}{\sqrt{k-i}} \leqq 4 \sqrt{\frac{M_j l_j}{k}},$$

and using this we conclude that

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \geqq \frac{\epsilon}{16} \sum_j \sqrt{l_j/M_j}.$$

5. The nonsimplicity of the fine structure of the coin-tossing game. The purpose of this section is to prove the following theorem.

THEOREM 5. *Let $x_0, x_1, \cdots$ be the successive times of return to equilibrium in the fair coin-tossing game. Then there exists no weighting $\{\lambda_m\}$, $\lambda_m \geqq 0$ of the positive even integers such that*

$$P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m \varepsilon A} \lambda_m < \infty.$$

PROOF. Consider any set $\bigcup_j I_j$ where the I_j are disjoint finite intervals which we can represent as $[m_j, m_j(1 + \alpha_j)]$, $0 < \alpha_j \leqq 1$, where $m_{j+1} \geqq 3m_j$. By theorem 4

$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0 \Leftrightarrow \sum_j \sqrt{\alpha_j} < \infty.$$

Let now $\{\lambda_m\}$ be any weighting of the positive even integers having the property stated in the theorem. By this property, $\lim_m \lambda_m = 0$ since otherwise we could find an indefinitely sparse set A which would be entered i.o. with probability one. We define a function $\phi(\alpha)$, $0 \leqq \alpha < \infty$ by

$$\phi(\alpha) = \liminf_n \sum_{m=n}^{ne^\alpha} \lambda_m,$$

where in writing the upper limit of summation as ne^α it is immaterial whether we take the next greater integer, or the previous integer.

PROPOSITION. *$\phi(\alpha)$ is monotone nondecreasing, $\phi(\alpha + \beta) \geqq \phi(\alpha) + \phi(\beta)$ and there is a neighborhood of the origin in which $\phi(\alpha) < \infty$.*

PROOF. The first assertion is immediate. As to the second, we write:

$$\liminf_n \left(\sum_{m=n}^{ne^\alpha e^\beta} \lambda_m\right) = \liminf_n \left(\sum_{m=n}^{ne^\alpha} \lambda_m + \sum_{m=ne^\alpha}^{ne^\alpha e^\beta} \lambda_m\right)$$

$$\geqq \liminf_n \left(\sum_{m=n}^{ne^\alpha} \lambda_m\right) + \liminf_n \left(\sum_{m=n}^{ne^\beta} \lambda_m\right).$$

Finally, suppose that $\phi(\alpha) = \infty$ for all $\alpha > 0$, and consider any sequence $\{\alpha_j\}$, $0 < \alpha_j \leqq 1$, such that $\sum_j \sqrt{\alpha_j} < \infty$. Since $\lim_n \sum_{m=n}^{n(1+\alpha_j)} \lambda_m = \infty$ for all j, we find a sequence of intervals $I_j = [m_j, m_j(1 + \alpha_j)]$ as far apart as desired having the undesirable property $\sum_j \sum_{m \varepsilon I_j} \lambda_m = \infty$.

To complete the proof of the theorem, we note that as a well-known consequence of the proposition there is a neighborhood N of the origin and a constant $q < \infty$ such that $\phi(\alpha) \leqq q\alpha$, $\alpha \varepsilon N$. Take $\{\alpha_j\}$, $\alpha_j > 0$, such that $\sum_j \alpha_j < \infty$ but $\sum_j \sqrt{\alpha_j} = \infty$, and $\{\alpha_j\} \subset N$. Then we may find a sequence $\{m_j\}$ increasing as rapidly as desired such that

$$\sum_{m=m_j}^{m_j(1+\alpha_j)} \lambda_m \leqq 2q\alpha_j.$$

Hence, taking $I_j = [m_j, m_j(1 + \alpha_j)]$ we have

$$\sum_j \sqrt{\alpha_j} = \infty \qquad \text{but} \qquad \sum_j \sum_{m \varepsilon I_j} \lambda_m < \infty.$$

It is a pleasure to acknowledge my debt to David Blackwell who brought my attention to the problems treated above.

REFERENCES

[1] D. Blackwell, "On transient Markov processes with a countable number of states and stationary transition probabilities," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 654-658.

[2] K. L. Chung and C. Derman, "Non-recurrent random walks," *Pacific J. Math.*, Vol. 6 (1956), pp. 441-447.

[3] L. Breiman, "On transient Markov chains with application to the uniqueness problem for Markov processes," *Ann. Math. Stat.*, to be published.

[4] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, 1953.

[5] K. L. Chung and P. Erdös, "On the application of the Borel-Cantelli lemma," *Trans. Amer. Math. Soc.*, Vol. 72 (1952), pp. 179-186.

A CONSTRUCTION FOR ROOM'S SQUARES AND AN APPLICATION IN EXPERIMENTAL DESIGN

By J. W. Archbold and N. L. Johnson

University College, London

1. T. G. Room [1] recently proposed the following problem: To arrange the $n(2n - 1)$ symbols rs (which is the same as sr) formed from all pairs of $2n$ different digits in a square of $2n - 1$ rows and columns such that in each row and column there appear n symbols (and $n - 1$ blanks) which among them contain all $2n$ digits.

He remarked that the problem is soluble when $n = 1$ (trivially) and $n = 4$ but not when $n = 2$ or 3; and he gave one solution for $n = 4$.

Squares of such a type have uses in experimental designs. We explain below a simple construction for squares where n has the form 2^{2m-1}. Each square constructed in this way is represented in a canonical form by applying a well-known theorem of J. Singer [2]. In this form as soon as the top row of entries in a square is known, all the other entries may be written down immediately by means of a straight-forward cyclic process. Thus an index of first rows is all that is necessary to catalogue squares in their canonical forms.

It may be permissible to give here a slight modification of the proof of Singer's theorem in order to show a natural application of the regular representation of linear algebras.

2. Let $\mathfrak{A}$ be a linear associative algebra, of order m and with modulus, over a commutative field K. It is well known that $\mathfrak{A}$ is isomorphic with an algebra of $m \times m$ matrices whose elements belong to K (c.f. Macduffee [3], Section 123).

A Galois field $GF(p^{mn})$ is such a linear algebra over a $GF(p^n)$. If the elements of the $GF(p^{mn})$ are $0, \alpha, \alpha^2, \cdots, \alpha^{p^{mn}-1} = 1$ the irreducible equation, of degree m and with coefficients in $GF(p^n)$,

$$f(x) \equiv x^m - a_1x^{m-1} - \cdots - a_m = 0$$

which is satisfied by α is called primitive (Dickson [4], Section 35). A basis for the algebra consists of $1, \alpha, \alpha^2, \cdots, \alpha^{m-1}$ and the modulus is 1.

The primitive equation is both the minimum and characteristic equation of the companion matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ a_m & a_{m-1} & a_{m-2} & \cdots & a_1 \end{pmatrix}$$

Received December 20, 1956; revised October 16, 1957.

The correspondence $\alpha^r \leftrightarrow \mathbf{A}^r$ determines an isomorphy, or regular representation, of the $GF(p^{mn})$ on the algebra or Galois field whose elements are the $m \times m$ matrices $0, \mathbf{A}, \mathbf{A}^2, \cdots, \mathbf{A}^{p^{mn}-1} = \mathbf{I}$, where $\mathbf{I}$ is the unit matrix (c.f. Macduffee [3], Section 109). If $N = 1 + p^n + \cdots + p^{(m-1)n}$ then the matrices $\mathbf{A}^{jN}$, for $j = 1, \cdots, p^n - 1$, are the multiples of $\mathbf{I}$ by the elements of $GF(p^n)$ and form a sub-algebra, of matrices, isomorphic with $GF(p^n)$.

In a finite projective space $PG(m - 1, p^n)$ over the $GF(p^n)$, let $\mathbf{x}$ and $\mathbf{y}$ denote column coordinate vectors. Then *the equation* $k\mathbf{y} = \mathbf{Ax}$, *where k is any non-zero element of $GF(p^n)$, determines a homography in the space of period N.* This is Singer's theorem; and the proof differs from his more in form than substance. It is significant for us that N is also the number of points in the space.

3. Confine attention now to the case where $p = 2$ and $n = 1$. The space, a $PG(m - 1, 2)$, contains $\mu, = 2^m - 1$, points, with three on every line.

The following are primitive irreducible polynomials over $GF(2)$:

$$x^2 - (x + 1), \quad x^3 - (x + 1), \quad x^4 - (x + 1),$$
$$x^5 - (x^2 + 1), \quad x^6 - (x + 1), \quad x^7 - (x + 1)$$
$$x^8 - (x^4 + x^3 + x^2 + 1), \quad x^9 - (x^8 + x^4 + x^3 + x^2 + 1).$$

This list is taken from Dickson ([4], p. 44); it is not exhaustive for the degrees mentioned but for each degree the second largest exponent of x is as small as possible.

For a given m, choose any appropriate primitive polynomial and consider the associated homography of $PG(m - 1, 2)$ of period μ. If P_1 is any point of the space, let its successive transforms under the homography be $P_2, P_3, \cdots, P_\mu$ $(P_{\mu+1} = P_1)$.

Now consider the space $PG(m - 1, 2)$ as being a prime in a $PG(m, 2)$. To achieve this, suppose $\mathbf{x}_1, \cdots, \mathbf{x}_\mu$ are coordinate vectors for $P_1, \cdots, P_\mu$. Then coordinate vectors for all but one of the points in $PG(m, 2)$ are obtained by adding a further zero or unit coordinate at the end of each $\mathbf{x}_i$; and the last point by taking coordinates consisting of m zeros followed by 1. Denote this last point by Q_0 and let Q_i be the third point on the line Q_0P_i; Q_i and P_i have the same first m coordinates.

To fix ideas, take $m = 3$ and $f(x) = x^3 - x - 1$. Then $\mu = 7$ and the corresponding homography is

$$k\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

or

$$y_0 : y_1 : y_2 = x_1 : x_2 : x_0 + x_1 .$$

Starting with $x_0 = 1, x_1 = x_2 = 0$, we obtain for $PG(3, 2)$ the following points:

$$\begin{array}{ccccccc} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \\ \begin{pmatrix}1\\0\\0\\0\end{pmatrix} & \begin{pmatrix}0\\0\\1\\0\end{pmatrix} & \begin{pmatrix}0\\1\\0\\0\end{pmatrix} & \begin{pmatrix}1\\0\\1\\0\end{pmatrix} & \begin{pmatrix}0\\1\\1\\0\end{pmatrix} & \begin{pmatrix}1\\1\\1\\0\end{pmatrix} & \begin{pmatrix}1\\1\\0\\0\end{pmatrix} \end{array}$$

$$\begin{array}{cccccccc} Q_0 & Q_1 & Q_2 & Q_3 & Q_4 & Q_5 & Q_6 & Q_7 \\ \begin{pmatrix}0\\0\\0\\1\end{pmatrix} & \begin{pmatrix}1\\0\\0\\1\end{pmatrix} & \begin{pmatrix}0\\0\\1\\1\end{pmatrix} & \begin{pmatrix}0\\1\\0\\1\end{pmatrix} & \begin{pmatrix}1\\0\\1\\1\end{pmatrix} & \begin{pmatrix}0\\1\\1\\1\end{pmatrix} & \begin{pmatrix}1\\1\\1\\1\end{pmatrix} & \begin{pmatrix}1\\1\\0\\1\end{pmatrix} \end{array}.$$

4. The idea is now to rename the points $Q_1, \cdots, Q_\mu$ as $R_1, \cdots, R_\mu$ in some order to be determined with the object, when possible, of ensuring that whenever the line Q_iQ_j passes through a point P_r then the line R_iR_j passes through a different point P_s.

The various incidences are then registered in a table of μ rows and μ columns as follows: if the line Q_iQ_j passes through P_r and R_iR_j passes through P_s make the entry

$$i, j \qquad (\text{or } j, i)$$

in the place belonging to the rth row and sth column of the table.

The number of entries in each row and column is the number of lines through a point of $PG(m, 2)$ which do not lie in a prime through the point. This number is $(2^m - 1) - (2^{m-1} - 1) = 2^{m-1}$. And the entries in every row and column are all the integers $0, 1, 2, 3, \cdots, 2^m - 1$ taken in pairs. No two pairs are the same and there are $2^{m-1}(2^m - 1)$ entries altogether.

In the cases examined below, the desired objective is reached when m is odd by defining R_t to be Q_u, where $u = 2^m - t$; and then no position in the incidence table contains more than one entry of the form (i, j). When m is even, the same definition is used for R_t but this leads to two entries in each position in the south-

TABLE 1

	1	2	3	4	5	6	7
1			24		56	37	01
2		35		67	41	02	
3	46		71	52	03		
4		12	63	04			57
5	23	74	05			61	
6	15	06			72		34
7	07			13		45	26

TABLE 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1					3, 9		12, 13			2, 5	7, 14		4, 15	8, 10 0, 2	0, 1 6, 11
2				4, 10		13, 14			3, 6	8, 15		5, 1	9, 11 0, 3	7, 12	
3			5, 11		14, 15			4, 7	9, 1		6, 2	10, 12 0, 4	8, 13		
4		6, 12		15, 1			5, 8	10, 2		7, 3	11, 13 0, 5	9, 14			
5	7, 13		1, 2			6, 9	11, 3		8, 4	12, 14 0, 6	10, 15				
6		2, 3			7, 10	12, 4		9, 5	13, 15 0, 7	11, 1					8, 14
7	3, 4			8, 11	13, 5		10, 6	14, 1 0, 8	12, 2					9, 15	
8			9, 12	14, 6		11, 7	15, 2 0, 9	13, 3					10, 1		4, 5
9		10, 13	15, 7		12, 8	1, 3 0, 10	14, 4					11, 2		5, 6	
10	11, 14	1, 8		13, 9	2, 4 0, 11	15, 5					12, 3		6, 7		
11	2, 9		14, 10	3, 5 0, 12	1, 6					13, 4		7, 8			12, 15
12		15, 11	4, 6 0, 13	2, 7					14, 5		8, 9			13, 1	3, 10
13	1, 12	5, 7 0, 14	3, 8					15, 6		9, 10			14, 2	4, 11	
14	6, 8 0, 15	4, 9					1, 7		10, 11			15, 3	5, 12		2, 13
15	5, 10					2, 8		11, 12			1, 4	6, 13		3, 14	7, 9

west to north-east diagonal of the table: no better definition for R_i has been devised which will prevent two entries from occurring in the same position.

5. For the case $m = 3$, which we began to consider in Section 3, let us define R_i to be Q_{8-i} ($i = 1, \cdots, 7$). We then obtain the incidences shown in Table 1.

It will be noticed that, beginning with the second, each row or column is obtained by a cyclic change in the positions, and values modulo 7, of the entries in the preceding row or column: that is, if X_{rs}, Y_{rs} are the entries in row r and column s and $X_{rs} \neq 0$, then, modulo 7,

$$X_{r,s} \equiv 1 + X_{r-1,s+1}, \qquad Y_{r,s} \equiv 1 + Y_{r-1,s+1}.$$

The whole table is therefore completely determined by the entries in any one row or column.

6. For $m = 4$, we have $\mu = 15$ and we take $f(x) = x^4 - x - 1$. R_i is now defined to be Q_{16-i} for $i = 1, \cdots, 15$. Table 2 is obtained.

Here the NE–SW diagonal is shared by two sets of entries. This is a characteristic feature arising when m is even but not when m is odd.

In fact, going now to the simplest case where $m = 2$ and $f(x) = x^2 - x - 1$, the table which arises is as follows:

	1	2	3
1			0 1 2 3
2		0 2 3 1	
3	0 3 1 2		

For $m = 5$, $\mu = 31$. Take $f(x) = x^5 - x^2 - 1$. Define R_i to be Q_{32-i}. Then we obtain Table 3 (only the first line of entries need be given).

TABLE 3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
12, 20	8, 23	9, 21	14, 15		10, 17		27, 29				2, 19			18, 31

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
22, 26	5, 11						16, 25	3, 6			13, 24		7, 28	4, 30	0, 1

7. If the columns of the first design of Section 5 be regarded as blocks, the rows as a set of treatments $a_1, \cdots, a_7$, and the numbers in the squares as a second set of treatments $b_0, \cdots b_7$, then the design is an incomplete block with respect to the first set of treatments and a complete randomized block with respect to the second set of treatments. The design is also balanced with respect to combinations of different levels of treatment a, with different levels of treatment b. The usual parametric model (Model I) would be

$$x_{tij} = A + B_t + \alpha_i + \beta_j + z_{tij}$$

(where x_{tij} denotes the observation on treatment combination $a_i b_j$ in the tth block, $\sum B_t = \sum \alpha_i = \sum \beta_j = 0$ and the z_{tij}'s are mutually independent random variables with common variance and mean zero). The analysis of variance appropriate to this model is obtained as follows:

(i) Carry out the standard incomplete block analysis on the means $\bar{x}_{ti}$ of pairs of observations for treatments a_i in the same (tth) block. Multiply the resultant sums of squares by two. This will give the Between Blocks and Adjusted Between Treatments a sums of squares in the final table.

(ii) Compute the Between Treatments b sum of squares in the usual way (that is, $7 \sum_{j=0}^{7} (\bar{x}_{..j} - \bar{x}_{...})^2$).

(iii) Compute the Residual sum of squares as Residual in (i) + $\sum_t \sum_i \sum_j (x_{tij} - \bar{x}_{ti.})^2$ − Between Treatment sum of squares in (ii).

The degrees of freedom appropriate to these sums of squares are then

Blocks	6
Adjusted Treatments a	6
Treatments b	7
Residual	36

One advantage of this design lies in the fact that the treatment b sum of squares is orthogonal to the treatment a sum of squares. It is, unfortunately, not possible to test for interaction between the two sets of treatments. Certain specific interactions may, however, be isolated from the Residual sum of squares. For example the contrast b_2 vs. b_4 in the presence of a_1 can be compared with the average effect of the same contrast in the presence of $a_2 a_3 \cdots a_7$, provided it is assumed that other interactions between a and b are negligible. The calculation of the sum of squares for such a contrast could be based on a two-way table with entries

$$b_2 a_1, \quad b_4 a_1 \qquad b_2 \sum_{i=2}^{7} a_i, \qquad b_4 \sum_{i=2}^{7} a_i$$

in the usual way.

Alternatively, the design may be regarded as an incomplete block design for treatments a, with main plots split for treatment b. In this case the design should be regarded as an incomplete block design also with respect to treatments b. The model becomes

$$x_{tij} = A + B_t + \alpha_i + \beta_j + u_{ti} + z_{tij}$$

where the u_{ti}'s are independent random variables, with zero mean and common variance, which are also independent of the z_{tij}'s. The two incomplete block analyses may be carried out separately (except that the Blocks sum of squares in the Treatments b analysis is the Total sum of squares in the Treatments a analysis). The sums of squares in the complete analysis, and their associated degrees of freedom, are

Blocks	6	
Adjusted Treatments a	6	As in the original
Error (i)	15	analysis (i)
Adjusted Treatments b	7	
Error (ii)	21	

As in the earlier analysis it is not possible, in general, to test for interaction between a and b, but certain specific interactions can be isolated from Error (ii).

Similar considerations apply to the second design of Section 7.

The design shown in paragraph 8 is a supplemented incomplete block design (in the sense of [5]) with respect to treatment a. The analysis of the design will, however, be similar to that described above for the designs of Section 7, and in particular the Treatment b sum of squares will again be orthogonal to the adjusted Treatment a sum of squares.

REFERENCES

[1] T. G. Room, "A new type of magic square," *Math. Gazette*, Vol. 39 (1955), p. 307.

[2] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. Amer. Math. Soc.*, Vol. 43, (1938), pp. 377–385.

[3] C. C. MacDuffee, *Introduction to Abstract Algebra*, John Wiley & Sons.

[4] L. E. Dickson, *Linear Groups*, Teubner, Leipzig (1901).

[5] T. N. Hoblyn, S. C. Pearce and G. H. Freeman, "Some considerations in the design of successive experiments in fruit plantations," *Biometrics*, Vol. 10, (1954), pp. 503–515.

A MULTIVARIATE TCHEBYCHEFF INEQUALITY[1]

BY INGRAM OLKIN[2] AND JOHN W. PRATT[3]

University of Chicago

0. Abstract. A multivariate Tchebycheff inequality is given, in terms of the covariances of the random variables in question, and it is shown that the inequality is sharp, i.e., the bound given can be achieved. This bound is obtained from the solution of a certain matrix equation and cannot be computed easily in general. Some properties of the solution are given, and the bound is given explicitly for some special cases. A less sharp but easily computed and useful bound is also given.

1. Introduction and outline. Tchebycheff's inequality states that if y is any real random variable with mean 0 and variance σ^2, then

$$P(|y| \geqq k\sigma) \leqq 1/k^2. \tag{1.1}$$

Berge [1] has generalized this result as follows. If y_1 and y_2 are any real random variables with means 0, variances σ_1^2 and σ_2^2 respectively, and correlation ρ, then

$$P(|y_1| \geqq k\sigma_1 \quad \text{or} \quad |y_2| \geqq k\sigma_2) \leqq \frac{1 + \sqrt{1-\rho^2}}{k^2}. \tag{1.2}$$

Berge gives an example where the inequality is achieved.

Suppose $y = (y_1, \cdots, y_p)$ is a random vector with mean 0 and nonsingular covariance matrix Σ. We seek an upper bound, depending on Σ and $k_1, \cdots, k_p$, for $P(|y_i| \geqq k_i\sigma_i$ for some $i)$.

The problem can be reduced by letting $x_i = y_i/(k_i\sigma_i)$. Then $x = (x_i, \cdots, x_p)$ has mean 0 and covariance matrix $\Pi = K^{-1}RK^{-1}$, where $R = (\rho_{ij})$ is the correlation matrix of y (and of x), $\Pi_{ij} = \sigma_{ij}/(\sigma_i\sigma_j k_i k_j) = \rho_{ij}/(k_i k_j)$, and K is a diagonal matrix with diagonal elements $k_1, \cdots, k_p$. Furthermore, $|y_i| \geqq k_i\sigma_i$ if and only if $|x_i| \geqq 1$, so $P(|y_i| \geqq k_i\sigma_i$ for some $i) = P(|x_i| \geqq 1$ for some $i)$.

Suppose A is a $p \times p$ matrix such that

$$xAx' \geqq 1 \quad \text{if} \quad |x_i| = 1 \quad \text{for some } i. \tag{1.3}$$

Then, looking at scalar multiples of x, we see that

$$xAx' \geqq 1 \quad \text{if} \quad |x_i| \geqq 1 \quad \text{for some } i, \tag{1.4}$$

Received February 18, 1957; revised June 11, 1957.

[1] This research was sponsored by the Office of Ordnance Research, U.S. Army, and the Statistics Branch, Office of Naval Research.

[2] Work done while on leave from Michigan State University.

[3] Now at Harvard University.

and that

$$xAx' \geqq 0 \quad \text{for all } x, \tag{1.5}$$

i.e., A is positive definite. Therefore

LEMMA 1.1. *If A satisfies* (1.3), *then*

$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) = P(|x_i| \geqq 1 \text{ for some } i) \leqq E(xAx') = \operatorname{tr} A\Pi, \tag{1.6}$$

where tr *denotes trace.*

Each A satisfying (1.3) therefore gives an upper bound for

$$P(|x_i| \geqq 1 \text{ for some } i).$$

The smallest bound obtainable in this way is the minimum of tr $A\Pi$ over all A satisfying (1.3). The set $\mathfrak{A}$ of all such matrices A is obviously convex, closed, and bounded from below, and tr $A\Pi$ is linear in A, so this minimum is achieved at an extreme point of $\mathfrak{A}$. In Theorem 3.3 it is shown that A is an extreme point of $\mathfrak{A}$ if and only if A^{-1} is positive definite and has 1's on the main diagonal. Furthermore, there is a unique extreme point of $\mathfrak{A}$ minimizing tr $A\Pi$, namely that extreme point A such that $A\Pi A$ is diagonal (Theorem 3.5). The bound thus obtained is the best possible, inasmuch as, if it is less than 1, there is a distribution for x (with mean 0 and covariance matrix Π) under which it is achieved, and otherwise there is a distribution for x under which

$$P(|x_i| \geqq 1 \text{ for some } i) = 1$$

(Theorem 3.7).

The minimizing matrix is easy to compute explicitly only in some special cases (Sec. 5). In the case $p = 2$, $k_1 = k_2 = k$, Berge lets $A = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}^{-1}$, shows that A satisfies (1.3), and minimizes tr $A\Pi$ with respect to a. Following this lead, in Sec. 2 we let $A = [(1 - a)I + ae'e]^{-1}$, where $e = (1, \cdots, 1)$, show that A satisfies (1.3) for $1 > a > -1/(p - 1)$, and minimize tr $A\Pi$ with respect to a, obtaining the bound in Theorem 2.3. Though the minimum over such A is in general, except in the case $p = 2$, not the minimum over all A satisfying (1.3), it provides a useful and easily computed bound. Lal [3] considers a matrix similar in form to that of Sec. 2. However, this does not lead to the best bound, as Lal asserts, and indeed his bound is not as tight as that given in Theorem 2.3 unless $p = 2$ or $\rho_{ij} = 0$ for all $i \neq j$.

2. A multivariate inequality. We will now carry out the program of the last paragraph.

LEMMA 2.1. $A = [(1 - a)I + ae'e]^{-1}$ *satisfies* (1.3) *if* $1 > a > -1/(p - 1)$.

PROOF. $A = [(1 - a)I + ae'e]^{-1} = (I - \alpha e'e)/(1 - a)$, *where* $\alpha = a/[1 + (p - 1)a]$. $x[I - \alpha e'e]x' = \sum x_i^2 - \alpha(\sum x_i)^2$

$$\geqq \begin{cases} \sum x_i^2 & \text{if } 0 \geqq a \geqq -1/(p-1), \quad \text{i.e.,} \quad \alpha \leqq 0; \\ (1 - p\alpha)\sum x_i^2 & \text{if } 0 \leqq a < 1, \quad \text{i.e.,} \quad 0 \leqq \alpha < 1/p \end{cases}$$

(The second case follows from $(\sum x_i)^2 \leqq p\sum x_i^2$.) The right-hand side becomes infinite with $\sum x_i^2$, so the minimum over all $(p - 1)$ – vectors z of

$$(1, z)(I - \alpha e'e)(1, z)'$$

occurs at a finite z. Differentiating

$$(1, z)(I - \alpha e'e)(1, z)' = 1 + \sum z_i^2 - \alpha(1 + \sum z_i)^2$$

with respect to each z_i we find that the minimizing z must satisfy $2z_i - 2\alpha(1 + \sum z_j) = 0$ for all i, or $z - \alpha z e'e - \alpha e = 0$. (Here e has $p - 1$ coordinates.) It follows that all z_i are equal, and that $\sum z_i = (p - 1)a$, so $z = ae$. Therefore the minimum over z of $(1, z)(I - \alpha e'e)(1, z)'$ is $1 - a$, and thus the minimum over z of

$$(1, z)A(1, z)'$$

is 1. The lemma follows. (See also Lemma 5.1.) || (This symbol will be used to indicate the end of a proof.)

LEMMA 2.2. *tr* $[(1 - a)I + ae'e]^{-1}\Pi$ *is minimized for* $1 > a > -1/(p - 1)$ *by*

$$a = \frac{t - \sqrt{u(pt - u)/(p - 1)}}{u - (p - 1)t}, \tag{2.1}$$

where $t = \operatorname{tr}\Pi = \sum \Pi_{ii} = \sum 1/k_i^2$ *and* $u = e\Pi e' = \sum \Pi_{ij} = \sum \rho_{ij}/(k_i k_j)$.

PROOF. $\operatorname{tr}[(1 - a)I + ae'e]^{-1}\Pi = \operatorname{tr}(I - \alpha e'e)\Pi/(1 - a) = (t - \alpha u)/(1 - a)$. The derivative of this quantity with respect to a has zeros at

$$a = \frac{t \pm \sqrt{u(pt - u)/(p - 1)}}{u - (p - 1)t}.$$

The condition $1 > a > -1/(p - 1)$ is satisfied if and only if

$$\mp\sqrt{u(pt - u)/(p - 1)}$$

is between $u/(p - 1)$ and $(pt - u)$. The upper sign is impossible because

$$u/(p - 1) \qquad \text{and} \qquad (pt - u)$$

are both positive. The lower sign is possible because $\sqrt{u(pt - u)/(p - 1)}$ is the geometric mean of $u/(p - 1)$ and $(pt - u)$. The extremum is a minimum since $(t - \alpha u)/(1 - a) \to \infty$ as $a \to 1$ or $a \to -1/(p - 1)$. ||

Substituting (2.1) in (1.6) and simplifying, we obtain, by Lemmas 1.1, 2.1, and 2.2, THEOREM 2.3. $P(|y_i| \geqq k_i\sigma_i \text{ for some } i) = P(|x_i| \geqq 1 \text{ for some } i)$

$$\leqq \frac{p-1}{p}t - \frac{p-2}{p^2}u + \frac{2}{p^2}\sqrt{u(pt-u)(p-1)}$$
$$= [\sqrt{u} + \sqrt{(pt-u)(p-1)}]^2/p^2.$$

In the case $p = 2$, we obtain

$$P(|y_1| \geqq k_1\sigma_1 \text{ or } |y_2| \geqq k_2\sigma_2) \leqq \frac{1}{2k_1^2k_2^2}[k_1^2 + k_2^2 + \sqrt{(k_1^2+k_2^2)^2 - 4\rho^2k_1^2k_2^2}],$$

which is Lal's equation (B), and is to be compared with Berge's result, (1.2).

3. The sharpest inequality. In this section we seek the tightest bound obtainable from Lemma 1.1, and show that it is sharp, following the outline in the next-to-last paragraph of Sec. 1. What we seek, then, is the minimum of tr $A\Pi$ for A satisfying (1.3), i.e., for $A \varepsilon \mathfrak{A}$. As remarked before, the minimum occurs at an extreme point of $\mathfrak{A}$. We start by characterizing, in Lemma 3.2, the matrices in $\mathfrak{A}$, and, in Theorem 3.3, the extreme points of $\mathfrak{A}$. We use the following lemma, which has some independent interest.

LEMMA 3.1. *If A is positive definite, the minimum of xAx' for $x_1 = 1$ is $1/b_{11}$ and occurs at $(1, b/b_{11})$, and only there, where*

$$B = \begin{pmatrix} b_{11} & b \\ b' & B_{22} \end{pmatrix} = A^{-1} = \begin{pmatrix} a_{11} & a \\ a' & A_{22} \end{pmatrix}^{-1}.$$

PROOF. It is easily checked that

$$b_{11} = (a_{11} - aA_{22}^{-1}a')^{-1}, \qquad b = -b_{11}aA_{22}^{-1}, \qquad B_{22} = A_{22}^{-1} + A_{22}^{-1}a'b_{11}aA_{22}^{-1}.$$

"Completing the square," we have

$$\begin{aligned}(1, z)A(1, z)' &= a_{11} + 2az' + zA_{22}z' \\ &= a_{11} - aA_{22}^{-1}a' + (z + aA_{22}^{-1})A_{22}(z + aA_{22}^{-1})' \\ &= b_{11}^{-1} + (z - b_{11}^{-1}b)A_{22}(z - b_{11}^{-1}b)'.\end{aligned}$$

Since A_{22} is positive definite, the lemma follows. Alternatively, $(1, z)A(1, z)'$ could be differentiated with respect to each coordinate of z, as in the proof of Lemma 2.1. ||

It follows from this lemma and (1.5) that

LEMMA 3.2. *$A \varepsilon \mathfrak{A}$ if and only if $B = A^{-1}$ is positive definite and $b_{ii} \leqq 1$, $i = 1, \cdots, p$.*

THEOREM 3.3. *A is extreme in $\mathfrak{A}$ if and only if $B = A^{-1}$ is positive definite and $b_{ii} = 1, i = 1, \cdots, p$.*

PROOF. (i) Suppose B is positive definite and all $b_{ii} = 1$. Then, by Lemma 3.2, $A \varepsilon \mathfrak{A}$. Suppose $A = (A_1 + A_2)/2$, $A_1 \varepsilon \mathfrak{A}$, $A_2 \varepsilon \mathfrak{A}$. For each i, by Lemma

3.1,

$$1 = 1/b_{ii} = \min_{x_i=1} xAx' \geqq \tfrac{1}{2}[\min_{x_i=1} xA_1 x' + \min_{x_i=1} xA_2 x'],$$

$$\min_{x_i=1} xA_1 x' \geqq 1. \qquad \min_{x_i=1} xA_2 x' \geqq 1.$$

It follows that

$$\min_{x_i=1} xA_1 x' = 1 = \min_{x_i=1} xA_2 x',$$

and the minima occur at the same point. This implies, by Lemma 3.1, that the ith row of A_1^{-1} equals the ith row of A_2^{-1}. As this is true for each i, $A_1 = A_2$. Therefore A is extreme in $\mathcal{A}$, which proves the "if".

(ii) If B is not positive definite, $A \varepsilon \mathcal{A}$, by Lemma 3.2. Suppose B is positive definite but $b_{ii} < 1$ for some i, say $b_{11} < 1$. Let

$$B(\delta) = B + \begin{pmatrix} \delta & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} b_{11} + \delta & b \\ b' & B_{22} \end{pmatrix}.$$

By Lemma 3.2, $B^{-1}(\delta) \varepsilon \mathcal{A}$ for δ small enough. If we can choose $\delta_1 \neq \delta_2$ such that $B^{-1}(\delta_1) \varepsilon \mathcal{A}$, $\quad B^{-1}(\delta_2) \varepsilon \mathcal{A}$, and

$$A = B^{-1} = \theta B^{-1}(\delta_1) + (1 - \theta)B^{-1}(\delta_2) \tag{3.1}$$

for some θ, $0 < \theta < 1$, we will have shown that A is not extreme in $\mathcal{A}$.

According to the first sentence of the proof of Lemma 3.1, with A and B interchanged, $B^{-1}(\delta)$ is a linear function of its upper left element $a_{11}(\delta)$, so (3.1) is equivalent to

$$a_{11} = a_{11}(0) = \theta a_{11}(\delta_1) + (1 - \theta)a_{11}(\delta_2).$$

Furthermore,

$$a_{11}(\delta) = \frac{1}{b_{11} + \delta - bB_{22}^{-1}b'} = \frac{1}{\delta + 1/a_{11}} = \frac{a_{11}}{1 + \delta a_{11}}.$$

Therefore (3.1) is equivalent to

$$\frac{\theta\delta_1}{1 + \delta_1 a_{11}} + \frac{(1 - \theta)\delta_2}{1 + \delta_2 a_{11}} = 0,$$

and it is clear that δ_1 and δ_2 can be chosen as desired. ||

This reduces the problem to that of minimizing $\operatorname{tr} B^{-1}\Pi$ for $B \varepsilon \mathcal{B}$, where $\mathcal{B}$ is the set of positive definite matrices with ones on the main diagonal. We will now show that $\operatorname{tr} B^{-1}\Pi$ is minimized at a unique interior point $\bar{B}$ of $\mathcal{B}$, (Theorem 3.4), and characterize $\bar{B}$ (Theorem 3.5).

THEOREM 3.4. $\operatorname{tr} B^{-1}\Pi$ *is a strictly convex function of* B *for* $B \varepsilon \mathcal{B}$, *and has a unique minimum, which occurs at an interior point* $\bar{B}$ *of* $\mathcal{B}$.

PROOF. Let $B(t)$ be a straight line in $\mathcal{B}$. Then dB/dt is a symmetric matrix,

$d^2B/dt^2 = 0$, and

$$\frac{d}{dt}\operatorname{tr} B^{-1}\Pi = -\operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\Pi,$$

$$\frac{d^2}{dt^2}\operatorname{tr} B^{-1}\Pi = 2\operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\Pi > 0.$$

This proves the strict convexity. The rest follows, since $\mathcal{B}$ is convex and bounded, and $\operatorname{tr} B^{-1}\Pi \to \infty$ as B approaches the boundary of $\mathcal{B}$. The latter follows from the fact that

$$\operatorname{tr} B^{-1}\Pi \geqq (\operatorname{tr} B^{-1})(\text{smallest eigenvalue of } \Pi). \quad \|$$

THEOREM 3.5. *$\bar{B}$ is the unique point of $\mathcal{B}$ such that $\bar{B}^{-1}\Pi\bar{B}^{-1}$, or equivalently $\bar{B}\Pi^{-1}\bar{B}$, is diagonal.*

PROOF. By Theorem 3.4, $\bar{B}$ is the unique point of $\mathcal{B}$ for which

$$\frac{d}{db_{ij}}\operatorname{tr} B^{-1}\Pi = \operatorname{tr} B^{-1}\left(\frac{dB}{db_{ij}}\right)B^{-1}\Pi = \operatorname{tr}\left(\frac{dB}{db_{ij}}\right)B^{-1}\Pi B^{-1} = 2c_{ij} = 0$$

for $i \neq j$, where $C = B^{-1}\Pi B^{-1}$, and dB/db_{ij} is a matrix with all elements zero except the (i, j)-th and (j, i)-th, which are one. $\|$

We note that $B^{-1}\Pi B^{-1} = C$ if and only if

$$B = \Pi^{1/2}(\Pi^{1/2}C\Pi^{1/2})^{-1/2}\Pi^{1/2} = C^{-1/2}(C^{1/2}\Pi C^{1/2})^{1/2}C^{-1/2}.$$

By Theorems 3.3, 3.4, and 3.5, the tightest inequality obtainable from Lemma 1.1 is

THEOREM 3.6. *$P(|y_i| \geqq k_i\sigma_i$ for some $i) = P(|x_i| \geqq 1$ for some $i)$*

$$\leqq \operatorname{tr}\bar{B}^{-1}\Pi = \operatorname{tr}\bar{B}^{-1}\Pi\bar{B}^{-1},$$

where $\bar{B}$ is the unique positive definite matrix having ones on the main diagonal such that $\bar{B}\Pi^{-1}\bar{B}$ is diagonal.

We note that $\operatorname{tr}\bar{B}^{-1}\Pi = \operatorname{tr}(\bar{B}^{-1}\Pi\bar{B}^{-1})\bar{B} = \operatorname{tr}\bar{B}^{-1}\Pi\bar{B}^{-1}$, since $\bar{B}^{-1}\Pi\bar{B}^{-1}$ is diagonal and $\bar{B}$ has ones on the diagonal.

According to the following theorem, the bound given in Theorem 3.6 is the smallest possible bound except when the smallest possible bound is the trivial bound 1.

THEOREM 3.7. *Let $\Theta = \bar{B}^{-1}\Pi\bar{B}^{-1}$ and $\theta_1, \cdots, \theta_p$ be its diagonal elements. Then*

$$\operatorname{tr}\bar{B}^{-1}\Pi = \operatorname{tr}\bar{B}^{-1}\Pi\bar{B}^{-1} = \operatorname{tr}\Theta = \sum \theta_i .$$

If $\sum \theta_i \leqq 1$, equality holds in Theorem 3.6 if and only if

$$P(x = b^i) = P(x = -b^i) = \theta_i/2, \qquad i = 1, \cdots, p,$$
$$P(x = 0) = 1 - \sum \theta_i , \tag{3.2}$$

where $b^1, \cdots, b^p$ *are the rows of* $\bar{B}$. *If* $\sum \theta_i > 1$, $P(|x_i| \geqq 1$ *for some* $i) = 1$ *if*

$$P(x = \sqrt{\textstyle\sum \theta_i}\, b^i) = P(x = -\sqrt{\textstyle\sum \theta_i}\, b^i) = \theta_i/(2 \textstyle\sum \theta_i), \qquad i = 1, \cdots, p. \tag{3.3}$$

Proof. If $\sum \theta_i \leqq 1$, (3.2) is a distribution for x, and if x has this distribution, equality holds in Theorem 3.6. If x has the distribution (3.3) and $\sum \theta_i > 1$, then, with probability one, $|x_i| \geqq \sqrt{\sum \theta_i} > 1$ for some i. In either case, x has mean 0 and covariance matrix

$$E(x'x) = \sum \theta_i b^{i\prime} b^i = \bar{B}\Theta\bar{B} = \Pi.$$

This proves the "if".

It remains to prove the "only if". Suppose $\sum \theta_i \leqq 1$ and equality holds in Theorem 3.6. Then, by the relation of (1.6) to (1.4) and (1.5), with probability one,

$$x\bar{B}^{-1}x' = 1 \quad \text{if} \quad |x_i| \geqq 1 \quad \text{for some } i,$$

and

$$x\bar{B}^{-1}x' = 0 \quad \text{otherwise}.$$

It follows, by Lemma 3.1, that the distribution of x is concentrated at 0 and $\pm b^1, \cdots, \pm b^p$. Then

$$E(x) = \sum [P(x = b^i) - P(x = -b^i)]b^i.$$

But $E(x) = 0$ and $b^1, \cdots, b^p$ are linearly independent, since they are the rows of a non-singular matrix, so $P(x = b^i) = P(x = -b^i)$ for all i. Then

$$E(x'x) = \sum 2P(x = b^i)b^{i\prime}b^i = \bar{B}D\bar{B},$$

where D is a diagonal matrix with diagonal elements

$$2P(x = b^1), \cdots, 2P(x = b^p).$$

But

$$E(x'x) = \Pi, \quad \text{so} \quad D = \bar{B}^{-1}\Pi\bar{B}^{-1} = \Theta,$$

and (3.2) follows.

4. On the solution of $\bar{B}\Theta\bar{B} = \Pi$. From $\Pi = \bar{B}\Theta\bar{B}$, we find that

$$\Pi_{ij} = \textstyle\sum_\alpha \bar{b}_{i\alpha}\theta_\alpha \bar{b}_{\alpha j},$$

and for $i = j$ we have the system of equations

$$1/k_i^2 = \sum_\alpha \bar{b}_{i\alpha}^2 \theta_\alpha, \qquad i = 1, \cdots, p.$$

If we write $\bar{B} \times \bar{B} = (\bar{b}_{ij}^2)$, then

$$(\theta_1, \cdots, \theta_p) = (k_1^{-2}, \cdots, k_p^{-2})(\bar{B} \times \bar{B})^{-1}.$$

Thus given $\bar{B}$ and $k_1, \cdots, k_p$, we can solve for Θ and Π. The matrix $B \times B$ is the Hadamard product, and is positive definite if B is ([2], p. 143). Given $k_1, \cdots, k_p$, $\bar{B}$ results from some Π if and only if $\bar{B} \in \mathcal{B}$ and

$$(k_1^{-2}, \cdots, k_p^{-2})(\bar{B} \times \bar{B})^{-1}$$

has positive elements. The following example shows that this last condition is not automatically satisfied.

$$B = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .5 \\ .8 & .5 & 1 \end{pmatrix}, \qquad |B \times B|(B \times B)^{-1} = \begin{pmatrix} .9375 & -.4800 & -.4800 \\ -.4800 & .5904 & .1596 \\ -.4800 & .1596 & .5904 \end{pmatrix},$$

$$k_1 = \cdots = k_p = 1.$$

Every $\bar{B} \in \mathcal{B}$ results from some $k_1, \cdots, k_p$ and Π, e.g., for

$$(k_1^{-2}, \cdots, k_p^{-2}) = (1, \cdots, 1)\bar{B} \times \bar{B}.$$

This section began with a procedure for determining Π from $\bar{B}$ by standard matrix operations. It appears that $\bar{B}$ cannot be obtained from Π by standard matrix operations except in special cases. We now give two properties of the solution (Theorems 4.1 and 4.2).

THEOREM 4.1. *If P is a permutation matrix and $P\Pi P = \Pi$, then $P\bar{B}P = \bar{B}$.*

PROOF.

$$(P\bar{B}P)\Pi^{-1}(P\bar{B}P) = P\bar{B}\Pi^{-1}\bar{B}P = P\Theta^{-1}P = \Theta^{-1}.$$

$P\bar{B}P \in \mathcal{B}$, so by the uniqueness in Theorem 3.5, $P\bar{B}P = \bar{B}$. ||

THEOREM 4.2. *If* $\Pi = \begin{pmatrix} \Pi_1 & 0 \\ 0 & \Pi_2 \end{pmatrix}$, *then* $\bar{B} = \begin{pmatrix} \bar{B}_1 & 0 \\ 0 & \bar{B}_2 \end{pmatrix}$, *where* $\bar{B}_i$ *minimizes*

$$\operatorname{tr} \bar{B}_i \Pi_i^{-1}, \qquad in \qquad \mathcal{B}_i, \qquad\qquad i = 1, 2.$$

PROOF. If $\bar{B}_i\Pi_i^{-1}\bar{B}_i$ is diagonal, $i = 1, 2$, then $\bar{B}\Pi^{-1}\bar{B}$ is diagonal, and by the uniqueness of $\bar{B}$, the conclusion follows. ||

5. Special cases.

THEOREM 5.1. *If $\Pi^{1/2}$ has equal diagonal elements, say, d, then*

$$\bar{B} = \Pi^{1/2}/d, \qquad \Theta = d^2 I$$

and

$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) = P(|x_i| \geqq 1 \text{ for some } i) \leqq \operatorname{tr} \bar{B}^{-1}\Pi = d^2 p.$$

This follows from Theorem 3.5. (The result for singular Π is an easy consequence of the result for non-singular Π.)

We note that $\Pi^{1/2}$ has equal diagonal elements if the group of permutation matrices P such that $P\Pi P = \Pi$ is transitive, i.e., every coordinate of x can be carried into every other one by a permutation of coordinates which preserves the covariances, i.e., $k_1 = \cdots = k_p$, and every coordinate of y can be carried into every other by a permutation of coordinates which preserves the correlations. This follows from the fact that $P\Pi^{1/2}P = \Pi^{1/2}$ if $P\Pi P = P$, since then $(P\Pi^{1/2}P)^2 = P\Pi P = \Pi$.

$\bar{B} = (1 - a)I + ae'e$, i.e., the inequality of Sec. 2 is the best possible, if and only if the elements of Π are

$$\Pi_{ii} = 1/k_i^2,$$

$$\Pi_{ij} = \rho_{ij}/k_i k_j = \frac{a}{1+a}\left[k_i^{-2} + k_j^{-2} + \frac{a(1-a)}{1+(p-1)a^2}\sum k_\alpha^{-2}\right], \tag{5.1}$$

in which case

$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) \leqq \operatorname{tr} \bar{B}^{-1}\Pi = \sum k_i^{-2}/[1 + (p-1)a^2]. \tag{5.2}$$

In the case $p = 2$, Π is always of this form and (5.2) yields (2.6).

If $k_1 = \cdots = k_p = k$, and $\Pi_{ii} = 1/k^2$, $\Pi_{ij} = \rho/k^2$, then Π is of the form (5.1) and

$$\begin{aligned} P(|y_i| \geqq k\sigma_i \text{ for some } i) &\leqq \operatorname{tr} \bar{B}^{-1}\Pi \\ &= \frac{p}{k^2[1+(p-1)a^2]} = \frac{[(p-1)\sqrt{1-\rho} + \sqrt{1+(p-1)\rho}]^2}{pk^2}. \end{aligned} \tag{5.3}$$

This could also be obtained from Theorem 2.3, or from Theorem 5.1.

$$\Pi^{1/2} = \frac{\sqrt{1-\rho}}{k} I + \frac{[\sqrt{1+p(-1)p} - \sqrt{1-\rho}]}{kp} e'e.$$

For special values of p and ρ we obtain in addition to Berge's result (1.2), the following inequalities.

(i) $\rho = 1$: $P(|y_i| \geqq k\sigma_i \text{ for some } i) \leqq 1/k^2$, which amounts to the univariate Tchebycheff inequality.

(ii) $\rho = 0$: For p *uncorrelated* random variables,

$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) \leqq \sum k_i^{-2},$$

whereas for p *independent* random variables, the univariate Tchebycheff inequality yields the bound $1 - \prod_{i=1}^p (1 - k_i^{-2})$.

(iii) $\rho = -1/(p-1)$: $P(|y_i| \leqq k\sigma_i \text{ for some } i) \leqq (p-1)/k^2$.

Acknowledgements. We wish to acknowledge the very helpful comments and suggestions of W. Hoeffding and L. J. Savage.

REFERENCES

[1] P. O. Berge, "A note on a form of Tchebycheff's theorem for two variables," *Biometrika*, Vol. 29 (1937), pp. 405–406.

[2] P. R. Halmos, "Finite dimensional vector spaces," *Annals of Mathematics Studies*, No. 7, Princeton University Press, Princeton, 1947.

[3] D. N. Lal, "A note on a form of Tchebycheff's inequality for two or more variables," *Sankhyā*, Vol. 15 (1955), pp. 317–320.

ON SELECTING A SUBSET WHICH CONTAINS ALL POPULATIONS BETTER THAN A STANDARD

By Shanti S. Gupta[1] and Milton Sobel

Bell Telephone Laboratories

1. Summary. A procedure is given for selecting a subset such that the probability that all the populations better than the standard are included in the subset is equal to or greater than a predetermined number P^*. Section 3 deals with the problem of the location parameter for the normal distribution with known and unknown variance. Section 4 deals with the scale parameter problem for the normal distribution with known and unknown mean as well as the chi-square distribution. Section 5 deals with binomial distributions where the parameter of interest is the probability of failure on a single trial. In each of the above cases the case of known standard and unknown standard are treated separately. Tables are available for some problems; in other problems transformations are used such that the given tables are again appropriate.

2. Introduction. C. W. Dunnett [3] has considered a different but related problem of comparing several treatment means with a control mean for normal distributions with a common unknown variance. His goal is to separate those treatments which are better than the control from those that are worse (or not better). He controls the probability of selecting the standard as the best (i.e., classifying all other treatments as worse) *when the treatments are all equal to (or worse than) the standard*. Earlier, E. Paulson [8] considered the problem of selecting the best one of k categories when comparing $k - 1$ categories with a standard. He deals with means of normal distributions with a common unknown variance and also with binomial distributions. He controls the probability of selecting the standard as the best *when the categories are equal to (or worse than) the standard*.

The procedure described in this paper controls the probability that the selected subset *contains all* those populations better than the control *for any possible true configuration*. If we define a correct decision as a selected subset which *contains all* those populations better than the standard, then the procedure given below guarantees a probability of a correct decision to be at least P^*, not only when the $k - 1$ populations are equal to (or worse than) the standard, but for any possible true configuration. Although we are comparing the procedure with the work noted above, it should be stressed that the goals are different and the procedures are not interchangeable. It should be noted that the treatment of Secs. 3 and 4 could be applied to several other distributions in the Koopman-Darmois family.

The goal treated in this paper is more flexible in that it allows the experimenter to choose a subset and withhold judgment about which is the best one. Then, if

Received March 14, 1957; revised July 1, 1957.

[1] Now at the University of Alberta.

the best one is desired it can be chosen from the selected subset on the basis of economic or other considerations.

Although the title and discussion above use the phraseology "populations better than a standard" we shall actually be interested in selecting all populations as good as or better than the standard; for practical purposes the distinction is of minor importance since in most of the practical problems the parameters of interest can have any value in some interval and are very rarely equal.

To discuss confidence statements we consider first the problems below in which the better populations are the ones with the larger values of the main parameter of interest τ. After the experiment is performed, we can make with confidence P^* the joint statement that *for all populations which are eliminated the parameter value is less than that of the standard.* This joint confidence statement follows from the fact that in selecting a subset *containing all* populations as good as or better than a standard we are automatically eliminating a subset containing only populations worse than the standard. Hence this procedure can be used *to eliminate* those populations which are distinctly inferior to the standard.

For the case in which the better populations are defined to be the ones with the smaller values of τ, the statistical problem is identical and all the results and tables of this paper apply with the obvious modifications.

3. Location parameter—normal populations. We shall assume that populations $\Pi_1, \Pi_2, \cdots, \Pi_p$ with unknown means $\mu_1, \mu_2, \cdots, \mu_p$, respectively are given and that Π_0 is the standard or control, whose mean μ_0 may or may not be known. For clarity we shall discuss the various cases separately.

Case A. Common known variance (μ_0 known). From each of the p populations $\Pi_i (i = 1, 2, \cdots, p)$, n_i independent observations are taken. Let $\bar{x}_i$ denote the sample mean from Π_i and let σ^2 be the common known variance.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\bar{x}_i \geqq \mu_0 - d\sigma/\sqrt{n_i}\text{."} \tag{3.1}$$

To determine the value of d let p_1, p_2 denote the true number of populations with $\mu \geqq \mu_0$ and $\mu < \mu_0$, respectively, so that $p_1 + p_2 = p$. Then the probability P of retaining all the p_1 populations with $\mu \geqq \mu_0$ is given by

$$\begin{aligned} P &= \prod_{i=1}^{p_1} P\{\bar{x}_i' \geqq \mu_0 - d\sigma/\sqrt{n_i'}\} \\ &= \prod_{i=1}^{p_1} P\{\sqrt{n_i'}(\bar{x}_i' - \mu_i')/\sigma \geqq -d + \sqrt{n_i'}(\mu_0 - \mu_i')/\sigma\}, \end{aligned} \tag{3.2}$$

where primes refer to values associated with the p_1 populations for which $\mu \geqq \mu_0$. Hence

$$P = \prod_{i=1}^{p_1} \{1 - F(-d + \sqrt{n_i'}(\mu_0 - \mu_i')/\sigma)\} \tag{3.3}$$

where $F(x)$ refers to the standard normal cumulative distribution function. The μ_i' in (3.3) are restricted by the condition $\mu_i' \geqq \mu_0$ and the minimum of (3.3) is attained by setting $\mu_i' = \mu_0 (i = 1, 2, \cdots, p_1)$. Since the result depends on the unknown integer p_1, we can obtain a lower bound by setting $p_1 = p$. Then using the symmetry of F we have

$$P \geqq F^p(d). \tag{3.4}$$

The equation determining d is obtained by setting the right-hand member of (3.4) equal to P^* and is given by

$$F(d) = (P^*)^{1/p}. \tag{3.5}$$

It should be noted that (3.4) is independent of μ_0, τ and n_i. Hence with a table of the standard normal c.d.f. one can easily find the appropriate d which satisfies (3.5) and is to be used in rule (3.1) for any μ_0, any σ and any vector n_i.

The case when the normal populations have different but known variances and the standard is known is treated similarly. The inequality defining the procedure for this problem, corresponding to (3.5), is

$$\bar{x}_i \geqq \mu_0 - d\sigma_i/\sqrt{n_i} \tag{3.6}$$

and the equation determining d is exactly the same as (3.5).

Case B. *Common known variance* (μ_0 *unknown*). In this case n_0 independent observations are taken on the standard Π_0. Let $\bar{x}_0$ denote the mean of these n_0 observations and let σ^2 be the known common variance for all the $(p + 1)$ populations. Then the procedure is to select all those populations for which the relation

$$\bar{x}_i \geqq \bar{x}_0 - d\sigma/\sqrt{n_i} \tag{3.7}$$

is satisfied. The equation determining d is obtained by the same argument as in Case A and, letting $f(x)$ denote the standard normal density, we obtain

$$\int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u \sqrt{\frac{n_i}{n_0}} + d \right) \right] f(u)\, du = P^*. \tag{3.8}$$

For the special case $n_i = n (i = 0, 1, \cdots, p)$ this reduces to

$$\int_{-\infty}^{\infty} F^p(u + d) f(u)\, du = P^*. \tag{3.9}$$

Equation (3.9) is independent of σ. Hence a single two-way table of d-values for different values of P^* and p solves the problem for all values of σ when $n_i = n(i = 0, 1, \cdots, p)$. Tables of d-values satisfying (3.9) for several values of P^* are given in [2] for $p = 1\ (1)\ 10$ and in [5] for $p = 1\ (1)\ 50$. A short table, using only two decimals of the original four, is excerpted from [5] (see Table I). In the more general case when the populations have different but known variances the procedure is defined by

$$\bar{x}_i \geqq \bar{x}_0 - d\sigma_i/\sqrt{n_i} \tag{3.10}$$

TABLE I[a]

Table of d-values satisfying (3.9) and used in the procedure defined by (3.7)

p	P^*			
	.75	.90	.95	.99
1	0.95	1.81	2.33	3.29
2	1.43	2.23	2.71	3.62
3	1.68	2.45	2.92	3.80
4	1.85	2.60	3.06	3.92
5	1.97	2.71	3.16	4.01
6	2.06	2.80	3.24	4.09
7	2.14	2.87	3.31	4.15
8	2.21	2.93	3.37	4.20
9	2.26	2.98	3.42	4.25
10	2.31	3.03	3.46	4.29
15	2.50	3.20	3.63	4.44
20	2.62	3.32	3.74	4.54
30	2.79	3.48	3.89	4.68
40	2.90	3.58	4.00	4.78
50	2.99	3.67	4.08	4.85

[a] For a more complete table see [5].

and the equation determining d is

$$\int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u \frac{\sigma_0}{\sigma_i} \sqrt{\frac{n_i}{n_0}} + d \right) \right] f(u)\, du = P^*. \tag{3.11}$$

this reduces to (3.9) in the case when $\sigma_i/\sqrt{n_i}$ = constant ($i = 0, 1, \cdots, p$).

Case C. *Common unknown variance* (μ_0 *known*). As in Case A, n_i observations are taken only on the p populations $\Pi_i (i = 1, 2, \cdots, p)$. Let s_ν^2 denote the pooled estimate of σ^2 based on $\nu = \sum_{i=1}^{p} (n_i - 1)$ degrees of freedom ($n_i > 1$ for at least one i). Then the procedure is to select those and only those populations Π_i for which

$$\bar{x}_i \geqq \mu_0 - ds_\nu/\sqrt{n_i}\,. \tag{3.12}$$

The equation determining d is

$$\int_0^{\infty} F^p(yd) q_\nu(y)\, dy = P^*, \tag{3.13}$$

where $q_\nu(y)$ is the density of $y = s_\nu/\sigma = \chi_\nu/\sqrt{\nu}$. This result holds for any μ_0 and depends on n_i only through the value of ν.

Case D. *Common unknown variance* (μ_0 *unknown*). In this case n_i observations are taken on all the populations $\Pi_i (i = 0, 1, \cdots, p)$ and the pooled estimate s_ν^2 of σ^2 is based on $\nu = \sum_{i=0}^{p} (n_i - 1)$ d.f. ($n_i > 1$ for at least one i).

The inequality defining the procedure is

$$\bar{x}_i \geqq \bar{x}_0 - ds_\nu/\sqrt{n_i}\,. \tag{3.14}$$

The equation determining d is

$$\int_0^\infty \int_{-\infty}^\infty \left[\prod_{i=1}^p F\left(u\sqrt{\frac{n_i}{n_0}} + yd\right)\right] f(u) q_\nu(y)\, du\, dy = P^*. \tag{3.15}$$

For $n_i = n (i = 0, 1, \cdots, p)$ this reduces to

$$\int_0^\infty \int_{-\infty}^\infty F^p(u + yd) f(u) q_\nu(y)\, du\, dy = P^*. \tag{3.16}$$

Methods for evaluating this double integral and tables of d-values for selected values of P^*, p and ν are given in [6] and values of $d/\sqrt{2}$ for other values of p and ν are given in [3].

4. Scale parameter—gamma or chi-square populations. In this section it will be more natural to define the population Π_i as better than Π_0 if the scale parameter $\theta_i < \theta_0$.

Case A. θ_0 *known.* We assume that the population $\Pi_i (i = 1, 2, \cdots, p)$ has the density

$$\frac{1}{\Gamma\left(\frac{\alpha_i}{2}\right)} \frac{1}{\theta_i^{-\alpha_i/2}} x^{\frac{\alpha_i}{2}-1} e^{-x/\theta_i}. \tag{4.1}$$

If $x_{ij} (j = 1, 2, \cdots, n_i)$ are the n_i observations on Π_i, then $t_i = \sum_{j=1}^{n_i} x_{ij}$ has the density (4.1) with α_i replaced by $\nu_i = n_i \alpha_i$ and the procedure is as follows.

Procedure: "Retain in the selected subset only those populations

$$\Pi_i (i = 1, 2, \cdots, p)$$

for which

$$\frac{t_i}{\nu_i} \leqq (1 + d)\theta_0." \tag{4.2}$$

Let q_1 and q_2 denote the number of populations with $\theta \leqq \theta_0$ and $\theta > \theta_0$, respectively, so that $q_1 + q_2 = p$. The probability P of a correct decision is given by

$$P = \prod_{i=1}^{q_1} P\left\{\frac{t_i'}{\theta_i'} \leqq (1 + d)\frac{\theta_0 \nu_i'}{\theta_i'}\right\}, \tag{4.3}$$

where primes refer to the q_1 populations with $\theta \leqq \theta_0$. Hence,

$$P = \prod_{i=1}^{q_1} G_{\nu_i'}\left[(1 + d)\frac{\theta_0 \nu_i'}{\theta_i'}\right], \tag{4.4}$$

where $G_{\nu_i}(x)$ is the c.d.f. of the gamma density in (4.1) with α_i replaced by ν_i and $\theta_i = 1$. A lower bound to this probability is obtained by setting $\theta_i = \theta_0$

and $q_1 = p$ so that the equation determining d can be written in the form

$$\prod_{i=1}^{p} \left\{ \frac{1}{\Gamma\left(\frac{\nu_i}{2}\right)} \int_0^{\nu_i(1+d)} u^{\frac{\nu_i}{2}-1} e^{-u}\, du \right\} = P^*. \tag{4.5}$$

For $\nu_i = \nu (i = 1, 2, \cdots, p)$ this is easily solved with the help of a table of the c.d.f. of $\gamma_\nu = \frac{1}{2}\chi_\nu^2$ with ν degrees of freedom.

Application to normal populations. If $\theta_i = 2\sigma_i^2 (i = 0, 1, \cdots, p)$ are the scale parameters for the $(p + 1)$ normal populations and $x_{ij} (j = 1, 2, \cdots, n_i)$ are the n_i observations on the population Π_i with the mean μ_i (known), then we retain the population Π_i in the selected subset if

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \leqq 2(1 + d)\sigma_0^2. \tag{4.6}$$

The equation determining the d in (4.6) is the same as (4.5) with ν_i replaced by n_i.

If the means μ_i are unknown and $n_i > 1 (i = 1, 2, \cdots, p)$, then in (4.6) we use the sample mean $\bar{x}_i$ in place of μ_i and $n_i - 1$ in place of n_i. The equation determining d is again (4.5) with $\nu_i = n_i - 1$.

Transformation: If we apply the transformation [1]

$$y_i = \ln\left(\frac{t_i}{\nu_i}\right) \qquad (i = 1, 2, \cdots, p), \tag{4.7}$$

then the procedure (4.2) of this section can be put in the form

$$y_i \leqq \ln\left(\frac{\theta_0}{2}\right) + d_1, \tag{4.8}$$

where

$$d_1 = \ln [2(1 + d)]. \tag{4.9}$$

Then using the normal approximation and the same argument as before, the approximate equation determining d_1 is

$$\prod_{i=1}^{p} \left\{ F\left(d_1 \sqrt{\frac{\nu_i}{2}}\right) \right\} = P^*. \tag{4.10}$$

For $\nu_i = \nu (i = 1, 2, \cdots, p)$ this gives an equation similar to (3.5). For the application to normal populations the equation corresponding to (4.8) is

$$\ln s_i^2 \leqq \ln \sigma_0^2 + d_1, \tag{4.11}$$

where d_1 is determined by (4.10) with $\nu_i = n_i$ or $n_i - 1$ according as the means μ_i are or are not known.

Case B. θ_0 *unknown.* The assumptions are the same as in Case A except that n_0 observations, viz., $x_{01}, x_{02}, \cdots, x_{0n_0}$ are taken on Π_0. The inequality defining the procedure and corresponding to (4.2) is

$$\frac{t_i}{\nu_i} \leqq (1 + d)\frac{t_0}{\nu_0}, \tag{4.12}$$

where $t_0 = \sum_{j=1}^{n_0} x_{0j}$ and $\nu_0 = n_0\alpha_0$. The equation determining d is obtained as before and is given by

$$\int_0^\infty \left[\prod_{i=1}^p \int_0^{\nu_i t(1+d)/\nu_0} \frac{u^{\frac{\nu_i}{2}-1} e^{-\frac{u}{2}}}{\Gamma(n_i/2)}\, du\right] \frac{t^{\frac{\nu_0}{2}-1} e^{-\frac{t}{2}}}{\Gamma(\nu_0/2)}\, dt. \tag{4.13}$$

Application to normal populations. For the case where the means are known, the rule takes the form

$$\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^2 \leqq \frac{(1 + d)}{n_0}\sum_{j=1}^{n_0}(x_{0j} - \mu_0)^2, \tag{4.14}$$

where d is given by (4.13) with $\nu_i = n_i$. If μ_i's are not known and

$$n_i > 1 (i = 0, 1, \cdots, p),$$

then the rule is the same as (4.14) with μ_i and n_i replaced by $\bar{x}_i$ and $n_i - 1$, respectively. The equation determining d is again (4.13) with $\nu_i = n_i - 1$.

Transformation: Using the transformation (4.7), we put the inequality defining the rule as

$$y_i \leqq y_0 + d_2. \tag{4.15}$$

The approximate equation determining d_2 is

$$\int_{-\infty}^\infty \left[\prod_{i=1}^p F\left(u\sqrt{\frac{n_i}{n_0}} + d_2\sqrt{\frac{n_i}{2}}\right)\right] f(u)\, du = P^*, \tag{4.16}$$

which for $n_i = n (i = 0, 1, \cdots, p)$ is of the same form as (3.9).

5. Binomial populations.

Case A. *Known standard.* It is assumed that $p + 1$ binomial populations Π_i with parameters $\theta_i (i = 0, 1, \cdots, p)$ are given where θ_0 is the known value of the probability of a unit being defective in the standard population Π_0. Again n_i independent observations are taken from each population

$$\Pi_i (i = 1, 2, \cdots, p).$$

Since θ_i is the probability of a unit being defective, we define Π_i to be better than Π_0 when $\theta_i < \theta_0$. Let x_i denote the number of defectives observed in the sample of n_i observations from $\Pi_i (i = 1, 2, \cdots, p)$.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\frac{1}{n_i} x_i \leqq \theta_0 + d\sqrt{\frac{\theta_0(1 - \theta_0)}{n_i}}. \text{''} \tag{5.1}$$

Let q_1, q_2, be defined as in Sec. 4; let $[m_i(d)]$ denote the largest integer in

$$m_i(d) = n_i\theta_0 + d\sqrt{n_i\theta_0(1 - \theta_0)} \qquad (i = 1, 2, \cdots, p). \tag{5.2}$$

The probability P of retaining all the q_1 populations with $\theta \leqq \theta_0$ is given by

$$P = \prod_{i=1}^{q_1}\left[\sum_{j=0}^{[m_i(d)]} C_j^{n_i}\theta_i^j(1 - \theta_i)^{n_i-j}\right]. \tag{5.3}$$

A lower bound is obtained by setting $\theta_i = \theta_0 (i = 1, 2, \cdots, p)$ and $q_1 = p$. The fact that $\theta_1 = \theta_0$ gives a lower bound can be shown by writing the binomial sum as an incomplete Beta function. Hence the inequality determining d becomes

$$\prod_{i=1}^{p}\left[\sum_{j=0}^{[m_i(d)]} C_j^{n_i}\theta_0^j(1 - \theta_0)^{n_i-j}\right] \geqq P^*, \tag{5.4}$$

and the solution is the smallest value of d satisfying (5.4). If $n_i = n$ then

$$[m_i(d)] = [m(d)]$$

and (5.4) reduces to

$$\sum_{j=0}^{[m(d)]} C_j^{n}\theta_0^j(1 - \theta_0)^{n-j} \geqq (P^*)^{1/p}. \tag{5.5}$$

This is easily solved by consulting a table of cumulative binomial probabilities.

For large values of n_i (large enough for the normal approximation to give good results) the inequality determining d can be approximated by the simple equation

$$F(d) = (P^*)^{1/p}, \tag{5.6}$$

where F is the standard normal c.d.f. This equation is independent of n_i and is much easier to solve than (5.4).

Case B. *Unknown standard.* The assumptions are the same as in Case A except that n_0 observations are taken on the standard population Π_0. Let x_0 be the number of defectives among n_0.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\frac{1}{n_i}x_i \leqq \frac{1}{n_0}x_0 + \frac{d}{2}\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}." \tag{5.7}$$

The probability P of retaining all the q_1 populations with $\theta \leqq \theta_0$ attains a minimum when $\theta_i = \theta (i = 0, 1, \cdots, p)$ and $q_1 = p$ and is given by

$$P(\theta, d) = \sum_{y=0}^{n_0}\prod_{i=1}^{p}\left[\sum_{j=0}^{[m_i(y,d)]} C_j^{n_i}\theta^j(1 - \theta)^{n_i-j}\right] C_y^{n_0}\theta^y(1 - \theta)^{n_0-y}, \tag{5.8}$$

where $[m_i(y, d)]$ is the largest integer contained in

$$m_i(y, d) = \frac{n_i}{n_0} y + \frac{dn_i}{2} \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}. \tag{5.9}$$

Then the desired value of d for (5.7) is the smallest number for which

$$\min_{0 \leq \theta \leq 1} P(\theta, d) \geqq P^*. \tag{5.10}$$

Since, except for very small n_i or very large p, the minimum occurs near $\theta = \frac{1}{2}$, we can obtain an approximate solution for d by finding the smallest number for which

$$P(\tfrac{1}{2}, d) \geqq P^*. \tag{5.11}$$

A simpler approximate solution, which gives good results when the n_i are not too small and p is not too large, is the normal approximation obtained under the assumption that $\theta_i = \frac{1}{2}(i = 0, 1, \cdots, p)$. Then from (5.7) we obtain for the approximate equation determining d

$$\int_{-\infty}^{\infty} \left[\prod_{i=1}^{p} F\left(u \sqrt{\frac{n_i}{n_0}} + d \sqrt{1+\frac{n_i}{n_0}} \right) \right] f(u)\, du = P^*. \tag{5.12}$$

For $n_i = n(i = 0, 1, \cdots, p)$ the rule (5.7) can be written as

$$x_i \leqq x_0 + d', \tag{5.13}$$

where $d' = d\sqrt{n/2}$. In carrying out the rule we can assume that d' is an *integer*. The desired value of d' is the smallest integer for which

$$\min_{0 \leq \theta \leq 1} \left\{ \sum_{y=0}^{n} \left[\sum_{j=0}^{y+d'} C_j^n \theta^j (1 - \theta)^{n-j} \right]^p C_y^n \theta^y (1 - \theta)^{n-y} \right\} \geqq P^*. \tag{5.14}$$

Then (5.12) can be written in the form

$$\int_{-\infty}^{\infty} F^p(u + \hat{d}) f(u)\, du = P^* \tag{5.15}$$

and the relation between d' and $\hat{d}$, using a continuity correction, is

$$d' = d\sqrt{\frac{n}{2}} = \left\{ \frac{\hat{d}\sqrt{n} - 1}{2} \right\} \tag{5.16}$$

where $\{x\}$ is the smallest integer greater than or equal to x.

Transformation: It may be desirable to solve the binomial problem by using an arc sine transformation and converting it into one involving the location parameter of the normal distributions. For example, for the Case B above with $n_i = n(i = 0, 1, \cdots, p)$ if we use the arc sine transformation as given in [4], the inequality defining the procedure is

$$\begin{aligned}\text{arc sin}\sqrt{\frac{x_i}{n+1}} + \text{arc sin}\sqrt{\frac{x_i+1}{n+1}} &\leqq \text{arc sin}\sqrt{\frac{x_0}{n+1}} \\ &+ \text{arc sin}\sqrt{\frac{x_0+1}{n+1}} + \frac{d\sqrt{2}}{\sqrt{2n+1}},\end{aligned} \tag{5.17}$$

where the approximate equation determining d is the same as (3.9) so that Table I is applicable here also.

REFERENCES

[1] M. S. BARTLETT AND D. G. KENDALL, "The statistical analysis of variance-heterogeneity and the logarithmic transformation," *J. Roy. Stat. Soc. Suppl.*, Vol. 8 (1946), pp. 128–138.

[2] R. E. BECHHOFER, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–29.

[3] C. W. DUNNETT, "A multiple comparison procedure for comparing several treatments with a control," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1096–1121.

[4] M. F. FREEMAN AND J. W. TUKEY, "Transformations related to the angular and the square root," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 607–611.

[5] S. S. GUPTA, "On a decision rule for a problem in ranking means," Mimeograph Series No. 150, May, 1956, Institute of Statistics, University of North Carolina, Chapel Hill, N. C.

[6] S. S. GUPTA AND M. SOBEL, "On a statistic which arises in selection and ranking problems," submitted for publication in *Ann. Math. Stat.*

[7] S. S. GUPTA AND M. SOBEL, "A procedure for selecting a subset containing the best of k binomial populations," to be published.

[8] E. PAULSON, "On the comparison of several experimental categories with a control," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 239–246.

SOME PROBLEMS OF SIMULTANEOUS MINIMAX ESTIMATION

By Stanisław Trybuła

Institute of Mathematics, Polish Academy of Sciences, Wroclaw

1. Summary. In this paper, we give minimax estimates of the parameters of the multivariate hypergeometric distribution and of the multinomial distribution, and of some parameters of an unspecified distribution with known range. We use as loss a weighted linear combination of squared differences between the true and the estimated values of the parameters. Some properties of the minimax estimates obtained are discussed.

2. Introduction. For our purpose, it is sufficient to define the estimation problem in a fixed sample size experiment as follows ([3], [4]). The random variable X is distributed in the space $\mathfrak{X}$ according to the distribution F belonging to the family $\mathfrak{F}$. We want to estimate $\omega(F)$ where ω is a function, the values of which belong to some space Ω, defined on $\mathfrak{F}$. (In the following we assume that X and $\omega(F)$ are vector valued.) An estimate is a statistic $f(X)$ having values in Ω. The nonnegative function $L[\omega(F), f(x)]$ is the loss resulting if, when F obtains, the estimate $f(x)$ is made. Define the risk by

$$R(f, F) = E\{L[\omega(F), f(X)] | F\} \tag{1}$$

and call $v(f) = \sup_{F\epsilon\mathfrak{F}} R(f, F)$ the guaranteed value for the estimate f. We seek the minimax estimate f^0, that is, the estimate whose guaranteed value is minimal. Obviously, such an estimate does not always exist. It is our aim to derive minimax estimates in some specific problems.

3. Problem 1. In practice, we often meet the following situation. A lot consisting of N units of a product has been produced. The units are classified into l categories, the ith category containing U_i units ($i = 1, \cdots, l$). A sample of size n is taken from the lot in which $k_1, \cdots, k_l$ units of categories $1, \cdots, l$ are observed. The problem is to estimate $U_1, \cdots, U_l$.

This leads to the estimation of the parameter $U = (U_1, \cdots, U_l)$ of a multivariate hypergeometric distribution. Thus, let

$$P(X_1 = k_1, \cdots, X_l = k_l) = \frac{\binom{U_1}{k_1} \cdots \binom{U_l}{k_l}}{\binom{N}{n}}. \tag{2}$$

It is known that,

$$m_i = E(X_i | U) = n\frac{U_i}{N}, \tag{3}$$

Received April 12, 1957; revised July 11, 1957.

$$\sigma_i^2 = E\{[X_i - E(X_i \mid U)]^2 \mid U\} = \frac{n(N-n)}{N^2(N-1)} U_i(N - U_i). \tag{4}$$

Suppose that the loss is

$$L(U,f) = \sum_{i=1}^{l} c_i[f_i(X) - U_i]^2 \qquad (c_i \geqq 0), \tag{5}$$

where $f = (f_1, \cdots, f_l)$ is the estimate of U and $X = (X_1, \cdots, X_n)$ is the sample. The risk is then

$$R(f, U) = E[L(U,f) \mid U] = E\left\{\sum_{i=1}^{l} c_i[f_i(X) - U_i]^2 \mid U\right\}. \tag{6}$$

If we study estimates of the form

$$f_i(X) = aX_i + b_i \qquad (i = 1, \cdots, l),$$

then

$$\begin{aligned} R(f, U) &= \sum_{i=1}^{l} c_i E\{[aX_i + b_i - U_i]^2 \mid U\} \\ &= \sum_{i=1}^{l} c_i[(am_i + b_i - U_i)^2 + a^2\sigma_i^2] \qquad (7) \\ &= \sum_{i=1}^{l} c_i\left[\left(a\frac{nU_i}{N} + b_i - U_i\right)^2 + a^2\frac{n(N-n)}{N^2(N-1)} U_i(N - U_i)\right]. \end{aligned}$$

Let the constant a assume a value such that the terms quadratic in U vanish. For this, it suffices to put

$$a = \frac{N}{n + \sqrt{n\dfrac{N-n}{N-1}}}.$$

If, moreover, we put

$$b_i = \frac{s_i N \sqrt{n\dfrac{N-n}{n-1}}}{n + \sqrt{n\dfrac{N-n}{N-1}}},$$

then (7) may be written

$$R(f, U) = \frac{nN\dfrac{N-n}{N-1}}{\left(n + \sqrt{n\dfrac{N-n}{N-1}}\right)^2} \sum_{i=1}^{l} c_i[Ns_i^2 + (1 - 2s_i)U_i]. \tag{8}$$

Without loss of generality, we may assume $c_1 \geqq c_2 \geqq \cdots \geqq c_l \geqq 0$. For the present, assume also that $c_2 \neq 0$. Let l_0 be the greatest index i for which $c_i \neq 0$

and let

$$L = \max_s \left[s \leqq l_0, \sum_{i=1}^{s} 1/c_i > \frac{s-2}{c_s} \right]. \tag{9}$$

The above assumptions being satisfied, we prove the following lemma:

If $L \leqq l$ then

$$\delta = \frac{L-2}{\sum_{j=1}^{L} 1/c_j} \geqq c_i \text{ for } i = L+1, L+2, \cdots, l. \tag{10}$$

PROOF. First, observe that a proof of the inequality is necessary only for $i = L + 1$. If $c_{L+1} = 0$, then the lemma obviously holds. If $c_{L+1} \neq 0$, it follows from the definition of L that

$$L - 1 \geqq c_{L+1} \sum_{j=1}^{L+1} \frac{1}{c_j} = 1 + c_{L+1} \sum_{j=1}^{L} 1/c_j .$$

The lemma is a direct consequence of this inequality.

Now put

$$s_i = \begin{cases} \dfrac{1}{2}\left(1 - \dfrac{\delta}{c_i}\right), & \text{when } i \leqq L, \\ 0, & \text{when } i > L. \end{cases} \tag{11}$$

Observe that $i \leqq L$, $0 < s_i \leqq \frac{1}{2}$. We shall show that the estimate

$$f^0 = (f_1^0, f_2^0, \cdots, f_l^0),$$

where

$$f_i^0(X) = N \frac{X_i + s_i \sqrt{n \dfrac{N-n}{N-1}}}{n + \sqrt{n \dfrac{N-n}{N-1}}}, \tag{12}$$

is the minimax estimate sought.

From (8) and (11) we have

$$R(f^0, U) = \frac{Nn \dfrac{N-n}{N-1}}{\left(n + \sqrt{n \dfrac{N-n}{N-1}}\right)^2} \left\{ \sum_{i=1}^{L} \left[c_i \frac{N}{4} \left(1 - \frac{\delta}{c_i}\right)^2 + \delta U_i \right] + \sum_{i=L+1}^{l} c_i U_i \right\}. \tag{13}$$

Observe that for

$$U_{L+1} = U_{L+2} = \cdots = U_l = 0, \tag{14}$$

$R(f^0, U) = c$, where c is a constant. By the lemma, $R(f^0, U) \leqq c$. Thus, by theorem 2.1 of [4], it is sufficient to prove that a distribution of the random variable U exists which satisfies (14) and for which f^0 is the Bayes estimate.

We seek for such a distribution among those of the form

$$P(U_{L+1} = \cdots = U_l = 0) = 1 \tag{15}$$

$$P(U_1 = u_1, \cdots, U_L = u_L) = C\,\frac{\Gamma(a_1 + u_1)\cdots\Gamma(a_L + u_L)}{u_1!\cdots u_L!}. \tag{16}$$

Let

$$r(f, P) = E[R(f, U)] = \sum_{i=1}^{l} c_i E\{E[f_i(X) - U_i]^2 \mid U\}. \tag{17}$$

It follows from (15) that the expected risk does not depend on f_i if $k_j \neq 0$ for at least one $j > L$. Thus, any estimate which minimizes (17) throughout the region $k_{L+1} = k_{L+2} = \cdots = k_l = 0$ is a Bayes estimate. Now, if

$$k_{L+1} = k_{L+2} = \cdots = k_l = 0$$

then, as is well-known, the expression (17) attains its minimum value for

$$\begin{aligned} &f_i(k_1, \cdots, k_L, 0, \cdots, 0) \\ &= E(U_i \mid X_1 = k_1, \cdots, X_L = k_L; X_{L+1} = \cdots = X_l = 0) \\ &= \begin{cases} 0 & \text{for } i > L; \\ \dfrac{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} u_i \prod_{j=1}^{L} \binom{u_j}{k_j} \frac{\Gamma(a_j + u_j)}{u_j!}}{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} \prod_{j=1}^{L} \binom{u_j}{k_j} \frac{\Gamma(a_j + u_j)}{u_j!}} & \text{otherwise.} \end{cases} \end{aligned} \tag{18}$$

The second part of Eq. (18) reduces to

$$\begin{aligned} &\frac{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} u_i \prod_{j=1}^{L} \frac{\Gamma(a_j + u_j)}{(u_j - k_j)!}}{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} \prod_{j=1}^{L} \frac{\Gamma(a_j + u_j)}{(u_j - k_j)!}} \\ &= \frac{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} [(a_i + k_i + v_i) - a_i] \prod_{j=1}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} \prod_{j=1}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}} \\ &= \frac{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} \Gamma(a_i + k_i + v_i + 1) \prod_{\substack{j=1 \\ j\neq i}}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} \prod_{j=1}^{L} \frac{\Gamma(a_i + k_j + v_j)}{v_j!}} - a_i \\ &= \frac{L_i}{M_i} - a_i. \end{aligned}$$

Observe that

$$
\begin{aligned}
&\sum_{\substack{v_1+\cdots+v_L=N-n\\ v_1\geq 0,\cdots,v_L\geq 0}} \frac{(N-n)!}{v_1!\cdots v_L!} \frac{\Gamma(b_1+v_1)\cdots\Gamma(b_L+v_L)}{\Gamma\left(N-n+\sum_{j=1}^{L} b_j\right)} \\
(19)\quad &= \int\cdots\int_{\substack{p_1+\cdots+p_L=1\\ p_1\geq 0,\cdots,p_L\geq 0}} p_1^{b_1-1}\cdots p_L^{b_L-1} \sum_{\substack{v_1+\cdots+v_L=N-n\\ v_1\geq 0,\cdots,v_L\geq 0}} \frac{(N-n)!}{v_1!\cdots v_L!} \\
&\qquad \cdot p_1^{v_1}\cdots p_L^{v_L}\,dp_1\cdots dp_L \\
&= \int\cdots\int_{\substack{p_1+\cdots+p_L=1\\ p_1\geq 0,\cdots,p_L\geq 0}} p_1^{b_1-1}\cdots p_L^{b_L-1}\,dp_1\cdots dp_L = \frac{\Gamma(b_1)\cdots\Gamma(b_L)}{\Gamma\left(\sum_{i=1}^{L} b_i\right)}.
\end{aligned}
$$

Applying (19) to L_i with $b_j = a_j + k_j$ for $j \neq i$ and

$$b_i = a_i + k_i + 1 (i = 1, \cdots, L)$$

and to M_i with $b_j = a_j + k_j$ we obtain

$$
\begin{aligned}
&\frac{L_i}{M_i} - a_i = \frac{(a_i+k_i)\left(N+\sum_{j=1}^{L} a_j\right)}{n+\sum_{j=1}^{L} a_j} - a_i \\
(20)\quad & \\
&= \frac{\left(N+\sum_{j=1}^{L} a_j\right)k_i + (N-n)a_i}{n+\sum_{j=1}^{L} a_j} = f_i^0(k_1, \cdots, k_L, 0, \cdots, 0)
\end{aligned}
$$

for

$$a_i = s_i \frac{N\sqrt{n\dfrac{N-n}{N-1}}}{N-n-\sqrt{n\dfrac{N-n}{N-1}}}.$$

Thus f^0 is minimax whenever $a_i > 0$; that is, when $N > n + 1$. For $N = n$ this result is immediate. For $N = n + 1$ it is a consequence of the fact that f^0 is Bayes for the a priori distribution of U defined by

$$P(U_{L+1} = \cdots = U_l = 0) = 1,$$

$$P(U_1 = u_1, \cdots, U_L = u_L) = \frac{N!}{u_1!\cdots u_L!} s_1^{u_1}\cdots s_L^{u_L}.$$

Up to this point, we have assumed $c_2 \neq 0$. Consider now the remaining cases. If all $c_i = 0$ then, obviously, every estimate is minimax. If alone $c_1 \neq 0$ then the problem may be considered as that of finding a minimax estimate of the parameter U in a one-dimensional hypergeometric distribution for the loss $L = (f - U)^2$. In this case, the formula for the minimax estimate is (see [4])

$$f(X) = N \frac{X + \frac{1}{2}\sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}. \tag{21}$$

It is easy to verify that the estimate (12) satisfies the condition

$$\sum_{i=1}^{L} f_i^0 = N.$$

Observe that we are actually dealing with only $l - 1$ independent parameters since one parameter, say U_l, may be computed from

$$U_1 + \cdots + U_l = N. \tag{22}$$

If we consider the problem of finding a minimax estimate for $U_1, \cdots, U_{l-1}$ under the loss

$$\bar{L}(U, f) = \sum_{i=1}^{l-1} \bar{c}_i(f_i - U_i), \tag{23}$$

the same estimate as above for $U_1, \cdots, U_{l-1}$ results as is seen by identifying c_i in the above with $\bar{c}_i (i = 1, \cdots, l - 1)$ and putting $c_l = 0$.

In solving our problem, we have restricted ourselves to the case $c_i \geqq 0$. If, however, some $c_j < 0$ then for $f_j \to \pm\infty$ the loss tends to $-\infty$ and, consequently, the problem becomes trivial.

In the special case $c_1 = c_2 = \cdots = c_l > 0$, formula (12) takes the form

$$f_i^0(x) = N \frac{X_i + \frac{1}{l}\sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}. \tag{24}$$

4. Corollaries for the multinomial case. For $N \to \infty$, the distribution of X converges to the multinomial distribution defined by

$$P(X_1 = k_1, \cdots, X_l = k_l) = \frac{n!}{k_1! \cdots k_l!} p_1^{k_1} \cdots p_l^{k_l},$$

$$0 \leqq k_i, 0 \leqq p_i \leqq 1, i = 1, \cdots, l'$$

$$\sum_{1}^{l} k_i = n;$$

and

$$\lim_{N\to\infty} \frac{f_i^0(X)}{N} = \frac{X_i + s_i\sqrt{n}}{n + \sqrt{n}} = g_i^0(x).$$

We shall prove[1] that $g^0 = (g_1^0, \cdots, g_l^0)$ is really a minimax estimate of the parameter $p = (p_1, \cdots, p_l)$ for the loss

$$L(g, p) = \sum_{i=1}^{l} c_i(g_i - p_i)^2, \qquad c_i \geqq 0. \tag{25}$$

When L, δ and s_i are defined by (9), (10), and (11), respectively, the loss is

$$R(g^0, p) = E\{L(g^0, p) \mid p\} = \frac{1}{(\sqrt{n} + 1)^2}\left[\sum_{i=1}^{L} (c_i s_i^2 + \delta p_i) + \sum_{i=L+1}^{l} c_i p_i\right], \tag{26}$$

which for $p_{L+1} = \cdots = p_l = 0$ is constant and, by the lemma of Sec. 2, maximum in p.

As is easy to verify, g^0 is Bayes for the a priori distribution $G(p)$ defined by

$$dG(p) = \begin{cases} C p_1^{\sqrt{n}s_1 - 1} \cdots p_L^{\sqrt{n}s_L - 1}, & \text{when } p_{L+1} = \cdots = p_l = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

By theorem 2.1 of [4] it follows that g^0 is minimax.

For $c_1 = \cdots = c_l > 0$, $s_i = 1/l$ and the minimax estimate g^0 takes the form

$$g_i^0(X) = \frac{X_i + \frac{1}{l}\sqrt{n}}{n + \sqrt{n}}.$$

This case was previously solved by H. Steinhaus in [5].

5. Problem 2. We shall prove the following theorem:

THEOREM. *Let X be a random variable distributed according to the unknown distribution F on the measurable space A. Let $g_1, \cdots, g_m$ be such bounded measurable functions on A that there exist two points x' x'' ε A such that each of these functions attains its minimum in x' and its maximum in x''. Let $X_1, \cdots, X_n$ be a random sample from F, and let $\lambda_i = E[g_i(X)]$. If the loss is given by*

$$L(f, \lambda) = \sum_{i=1}^{m} c_i(f_i - \lambda_i)^2, \tag{29}$$

where $f = (f_1, \cdots, f_m)$ is an estimate of $\lambda = (\lambda_1, \cdots, \lambda_m)$, then the minimax estimate of λ is given by

$$f_i^0(X_1, \cdots, X_n) = \frac{\sum_{j=1}^{n} g_i(X_j)}{n + \sqrt{n}} + \frac{s_i}{\sqrt{n} + 1} \tag{30}$$

(s_i is the arithmetic mean of the maximum and minimum values of $g_i(x)$).

PROOF. If $\bar{f}_i(X_1, \cdots, X_n) = a \sum_{j=1}^{n} g_i(X_j) + b_i$, then the risk may be

[1] While this paper was being written, Joseph Dubay communicated to me a result similar to this but not in its full generality.

written

$$(31)\quad \begin{aligned} R(\hat{f}, F) &= E\left[\sum_{i=1}^{m} c_i(f_i - \lambda_i)^2 \mid F\right] = \sum_{i=1}^{m} c_i E\left\{\left[a\sum_{j=1}^{n} g_i(X_j) + b_i - \lambda_i\right]^2 \middle| F\right\} \\ &= \sum_{i=1}^{m} c_i\{[(1 - an)\lambda_i - b_i]^2 + na^2E\{[g_i(X) - \lambda_i]^2 \mid F\}\}. \end{aligned}$$

Let

$$\alpha_i = \min_{x\epsilon A} g_i(x) = g_i(x'), \qquad \beta_i = \max_{x\epsilon A} g_i(x) = g_i(x'').$$

It is easy to prove that

$$(32)\qquad E\{[g_i(X) - \lambda_i]^2 \mid F\} \leqq (\beta_i - \lambda_i)(\lambda_i - \alpha_i).$$

Thus

$$(33)\qquad R(\hat{f}, F) \leqq \sum_{i=1}^{m} c_i\{[(1 - an)\lambda_i - b_i]^2 + na^2(\beta_i - \lambda_i)(\lambda_i - \alpha_i)\}.$$

Putting

$$a = \frac{1}{n + \sqrt{n}}, \qquad b_i = \frac{s_i}{\sqrt{n} + 1},$$

we obtain

$$(34)\qquad R(f^0, F) \leqq \frac{1}{4(\sqrt{n} + 1)^2} \sum_{i=1}^{m} c_i(\beta_i - \alpha_i)^2 = c.$$

Observe that if a distribution $\bar{F}$ of the random variable X is defined by

$$(35)\qquad \begin{aligned} P(X = x') &= 1 - p, \\ P(X = x'') &= p. \end{aligned}$$

Then $\lambda_i = \alpha_i + (\beta_i - \alpha_i)p$, and equality obtains in (32); i.e.

$$(36)\qquad R(f^0, \bar{F}) = c.$$

The distribution F depends on the parameter p. Since (34) and (36) hold, it is sufficient to show (as in Sec. 3) that there exists a distribution G of p for which (30) is Bayes—that is, a distribution G such that (30) minimizes the expected risk

$$\begin{aligned} r(f, G) = E[R(f, \bar{F})] &= \sum_{i=1}^{m} c_i\, E\{E[(\lambda_i - f_i)^2 \mid \bar{F}]\} \\ &= \sum_{i=1}^{m} c_i E\{E\{[\alpha_i + (\beta_i - \alpha_i)p - f_i]^2 \mid p\}\}. \end{aligned}$$

It is easy to verify that this happens for the distribution $G^0(p)$ defined by equation

$$dG^0(p) = C(pq)^{(\sqrt{n}/2)-1}\,dp \qquad (q = 1 - p). \tag{37}$$

This completes the proof.

6. In this paper, we have used the loss $L = \sum_{i=1}^{m} c_i(f_i - \omega_i)^2$. This loss has been extensively investigated ([2], [4], [5], [6]). For many special problems, other loss functions might be used, for example,

$$L = \sum_{i=1}^{m} c_i \mid f_i - \omega_i \mid,$$

about which little is known at present.

Problems considered in this paper were suggested to me by L. J. Savage and H. Steinhaus. I am indebted to J. A. Dubay and L. J. Savage for help and suggestions made during the preparation of this paper.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, Inc., New York, 1954.

[2] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.

[3] P. R. Halmos, "The theory of unbiased estimation," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 34–43.

[4] J. L. Hodges, Jr., and E. L. Lehmann, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[5] H. Steinhaus, "The problem of estimation," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 633–648.

[6] S. Trybuła, "On minimaksowej estymacji parametrów w rozkładzie wielomianowym," to be published.

[7] A. Wald, "Contributions to the theory of statistical estimation and testing hypothesis," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 229–325.

[8] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, Inc., New York, 1950.

A THEOREM ON FACTORIAL MOMENTS AND ITS APPLICATIONS

By P. V. Krishna Iyer

Defense Science Laboratory, New Delhi

0. Summary. The theorem that the sth factorial moment for the sum of N events is $s!$ times the sum of the expectations for any s of the events occurring simultaneously has been proved by induction. The applications of this result in obtaining easily the moments of a number of distributions arising from a sequence of observations belonging to two continuous populations and other cases have been demonstrated.

1. Introduction. A number of distributions arising from a sequence of n observations belonging to a binomial population have been considered by the writer [3], [4] in some of his earlier publications. The moments of these distributions were obtained by using the theorem that the sth factorial moment is equal to $s!$ times the expectation for s of the characters considered in the distribution. Thus for a sequence of observations consisting of A's and B's with the probabilities p and q respectively, the sth factorial moment for the distribution of the total number of AB and BA joins between successive observations is the expectation for s joins like AB and BA in the sequence. It can be seen that there are s different ways of obtaining s joins. They are:

(1) From $(s + 1)$ consecutive observations.

(2) From two sets of l_1 and l_2 consecutive observations such that $l_1 + l_2 - 2$ is equal to s.

(3) From three sets of l_1, l_2 and l_3 consecutive observations such that

$$l_1 + l_2 + l_3 - 3$$

is equal to s.

(4) From k sets of $l_1, l_2, \cdots, l_k$ consecutive observations subject to the condition

$$\sum_1^k l_r - k = s,$$

where k takes values 1 to s.

The sum of the expectations for $1, 2, 3, \cdots, s$ is equal to $1/s!$ (the sth factorial moment for the distribution of the total number of AB and BA joins of the sequence).

The theorem as it stands appears to be applicable only for the distributions arising from a binomial sequence consisting of A's and B's with fixed probabilities p and q. We shall show in this paper that this result can be applied for distributions arising from two samples belonging to populations with cumu-

Received December 19, 1956; revised May 29, 1957.

lative distribution functions F and G. Before discussing this aspect, we shall first give a rigorous proof of the theorem and then show how it can be applied for the case of continuous distributions. The use of the result for distributions arising from Markoff chain is also illustrated.

2. Statement of theorem and proof.

THEOREM. *The* sth *factorial moment about the origin of any statistic* X *which is the sum of* N *events, dependent or independent, is equal to* $s!$ *times the sum of the expectations for any* s *of the events occurring together.*

PROOF. Let the events be denoted by $x_1, x_2, \cdots, x_N$. As in the case of binomial distribution, assume that the x's take value 1 if the event occurs and zero otherwise. Define

$$X = \sum_1^N x_r,$$

$$(1) \qquad E(X) = E(\textstyle\sum x_r) = \sum E(x_r)$$

$$= \text{the sum of the expectations of the different events}$$

$$= \text{the sum of the probabilities for the events to occur.}$$

$$E(X^2) = E(\textstyle\sum x_r)^2 = E(\sum x_r^2) + 2E(\sum x_r x_s), \qquad s > r.$$

Now

$$E(\textstyle\sum x_r^2) = E(X);$$

hence

$$E(X^2) = E(X) + 2E(\textstyle\sum x_r x_s)$$

or

$$(2) \quad E\{X(X-1)\} = 2\textstyle\sum E(x_r x_s)$$

$$= 2 \text{ (sum of the expectations for any two of the events)}$$

$$= 2 \text{ (sum of the probabilities for any two of the events to occur together).}$$

$$E(X^3) = E(\textstyle\sum x_r)^3 = E(\sum x_r^3) + 3E\sum (x_r^2 x_s) + 3E(\sum x_r x_s^2) + 6E(\sum x_r x_s x_t), \qquad t > s > r$$

since

$$E(\textstyle\sum x_r^3) = E(X),$$

$$E(x_r^2 x_s) = E(x_r x_s),$$

$$E(x_r x_s^2) = E(x_r x_s),$$

$$E(X^3) = E(X) + 6E(\textstyle\sum x_r x_s) + 3!E(\sum x_r x_s x_t).$$

Substituting the value of $E(\sum x_r x_s)$ from (2), we get

$$E(X^3) = E(X) + 3E\{X(X-1)\} + 3!E(\sum x_r x_s x_t)$$

or

$$(3) \quad E\{X(X-1)(X-2)\} = 3!\sum E(x_r x_s x_t)$$

= 3!(sum of the expectations for any three of the events)

= 3!(sum of the probabilities for any three of the events to occur together).

Thus the theorem holds good for $s = 1$ to 3.

It may be noted that the results given above hold good even without taking the expectation of both sides because the x's take values 1 or 0 only.

We shall now establish the general relation by induction. For this we show that if

$$(4) \qquad X^{[s]} = s!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s})$$

holds good for any value of s, it is true for $(s + 1)$ also.

Multiplying both sides of (4) by X we get

$$\begin{aligned}[X^{[s]}X] &= s!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s})(\sum x_r) \\ &= (s+1)!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s} x_{t_{s+1}}) \\ &+ s!\,[\sum x_{t_1}^2 x_{t_2} \cdots x_{t_s} \\ &+ \sum x_{t_1} x_{t_2}^2 \cdots x_{t_s} + \cdots + \sum x_{t_1} x_{t_2} \cdots x_{t_s}^2] \\ &= (s+1)!\sum x_{t_1} x_{t_2} \cdots x_{t_{s+1}}) \\ &+ s!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s}).\end{aligned}$$

Substituting for $\sum x_{t_1} x_{t_2} \cdots x_{t_s}$ from (4), (5) reduces to

$$(6) \qquad X^{[s]}(X - s) = X^{[s+1]} = (s+1)!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_{s+1}}).$$

Taking the expectation of both sides

$$(7) \qquad E\{X^{[s+1]}\} = (s+1)!\sum E(x_{t_1} x_{t_2} \cdots x_{t_{s+1}}).$$

Hence the theorem.

3. Applications. We shall now examine how the above result can be applied for obtaining easily the moments of a number of distributions including those arising from a simple Markoff chain. Some of the distributions considered here have been discussed by Wald and Wolfowitz [7], Mood [6], Mann & Whitney [5], and others.

(1) *Binomial distribution*. It is obvious that the rth factorial moment for the distribution of x, the number of successes out of n trials is given by

$$\frac{\mu'_{[r]}}{r!} = \binom{n}{r} p^r, \tag{8}$$

where p is the probability for a success.

(2) *Hypergeometric distribution.* This can be deduced from the above by substituting

$$p^r = \frac{(N - M)^{[r]}}{N^{[r]}},$$

where N and M have the usual significance. This follows from the fact that the probability p for the 1*st*, 2*nd*, 3*rd*, $\cdots$ successes are

$$\frac{N - M}{N}, \quad \frac{N - M - 1}{N - 1}, \quad \frac{N - M - 2}{N - 2}, \cdots$$

(3) *Distribution of the number of AB joins between successive observations of a binomial sequence.* We first note that r AB joins can be formed from only r sets of two consecutive observations each and therefore

$$\frac{\mu'_{[r]}}{r!} = \binom{n - r}{r} p^r q^r. \tag{9}$$

This can be seen from the fact that the probabilities for an AB join is pq and that there are $\binom{n - r}{r}$ ways of obtaining them from n observations in a sequence.

(4) *Distribution of AB joins for binomial sequence of n_1A's and n_2B's.* As in the case of hypergeometric series, we substitute

$$p^r q^s = \frac{n_1^{[r]} n_2^{[s]}}{(n_1 + n_2)^{[r+s]}}$$

in the results given in (3) above. Thus

$$\mu'_{[r]} = \frac{(n_1 + n_2 - r)^{[r]} n_1^{[r]} n_2^{[r]}}{(n_1 + n_2)^{[2r]}}. \tag{10}$$

(5) *Distribution of AB and BA joins between consecutive observations of a binomial sequence.* Taking for simplicity the third factorial moment, we note that three joins can be obtained from (i) four consecutive observations $ABAB$ or $BABA$, (ii) two sets, one of two and the other of three consecutive observations like $AB - ABA$; $BA - ABA$; $AB - BAB$ and $BA - BAB$, (iii) three sets of each of two consecutive observations AB or BA. The sum of the expectations for the above three ways of obtaining three joins is

$$\frac{\mu'_{[3]}}{3!} = 2(n - 3)p^2q^2 + 8\binom{n - 3}{2} p^2q^2 + \binom{n - 3}{3} 8p^3q^3. \tag{11}$$

(6) *A sequence formed by pooling two samples A and B belonging to F.* Let two samples A and B of sizes n_1 and n_2 be drawn from a population where cumulative

distribution function is $F(x)$, $F(x)$ being continuous and x taking values from $-\infty$ to $+\infty$. By pooling together A and B and arranging them in ascending or descending order we obtain a sequence of A's and B's as in (4) considered above. Hence the moments of any distribution arising from this sequence can be obtained from the corresponding ones for the binomial sequence by substituting

$$p^r q^s = \frac{n_1^{[r]} n_2^{[s]}}{(n_1 + n_2)^{[r+s]}}.$$

(7) *Same as* (6), $F \neq G$. The calculation of the moments for some of the distributions discussed above is more complicated when $F \neq G$. We shall show below how the present theorem enables us to obtain the moments of these distributions also.

(a) *Number of observations of sample A to the left of the* rth *value of the combined ordered sequence of* A's *and* B's.

$$\frac{\mu'_{[s]}}{s!} = \text{[Number of ways of selecting } A\text{'s from } (r-1) \text{ observations]}$$
$$\times \text{[Probability that } s \text{ out of the } (r-1) \text{ values to the left of the } r\text{th observation belong to } A]$$

Assuming $n_1F(\alpha) + n_2G(\alpha) = r$, the probability that amongst the $(r-1)$ values to the left of the rth observation there are s A's is

$$\frac{n_1F(\alpha)}{n_1 F(\alpha) + n_2 G(\alpha)} \frac{(n_1 - 1)F(\alpha)}{(n_1 - 1)F(\alpha) + n_2 G(\alpha)} \frac{(n_1 - 2)F(\alpha)}{(n_1 - 2)F(\alpha) + n_2 G(\alpha)} \cdots s \text{ terms.}$$

Using the relation between $F(\alpha)$, $G(\alpha)$ and r we get

$$(12) \qquad \frac{\mu'_{[s]}}{s!} = \binom{r-1}{s} \frac{n_1^{[s]}[F(\alpha)]^s}{r[r - F(\alpha)][r - 2F(\alpha)] \cdots [r - (s-1)F(\alpha)]}.$$

(b) *Number of AB and BA joins between successive observations.* As the higher moments are complicated we shall be content to obtain the second moment.

$$\frac{\mu'_{[2]}}{2!} = \text{the sum of the expectations for two joins from (i) three consecutive observations and (ii) two sets each of two consecutive observations.}$$

Expectation for two joins from three consecutive observations x_1, x_2, and x_3 is given by

$$\begin{aligned} & n_1(n_1 - 1)n_2 \int_{-\infty}^{+\infty} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} \{1 - F(x_3) + F(x_1)\}^{n_1-2} \\ (13) \qquad & \cdot \{1 - G(x_3) + G(x_1)\}^{n_2-1} \, dF(x_1) \, dG(x_2) \, dF(x_3) \\ & + n_1 n_2(n_2 - 1) \int_{-\infty}^{+\infty} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} \{1 - G(x_3) + G(x_1)\}^{n_2-2} \\ & \cdot \{1 - F(x_3) + F(x_1)\}^{n_1-1} \, dG(x_1) \, dF(x_2) \, dG(x_3), \qquad x_1 < x_2 < x_3. \end{aligned}$$

Expectation for four joins from two sets of two consecutive observations each is equal to

$$
(14)\quad
\begin{aligned}
n_1^{[2]} n_2^{[2]} \Bigg[\int_{-\infty}^{+\infty} \int_{-\infty}^{x_4} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} & A \, dF(x_1) \, dG(x_2) \, dF(x_3) \, dG(x_4) \\
& + \int_{-\infty}^{+\infty} \int_{-\infty}^{x_4} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} A \, dF(x_1) \, dG(x_2) \, dG(x_3) \, dF(x_4) \\
& + \int_{-\infty}^{+\infty} \int_{-\infty}^{x_4} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} A \, dG(x_1) \, dF(x_2) \, dF(x_3) \, dG(x_4) \\
& + \int_{-\infty}^{+\infty} \int_{-\infty}^{x_4} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} A \, dG(x_1) \, dF(x_2) \, dG(x_3) \, dF(x_4),
\end{aligned}
$$

where

$$
\begin{aligned}
A = \{1 - F(x_2) + F(x_1) - F(x_4) + F(x_3)\}^{n_1 - 2} & \\
\times \{1 - G(x_2) + G(x_1) - G(x_4) + G(x_3)\}^{n_2 - 2}&, \\
x_1 < x_2 < x_3 < x_4&.
\end{aligned}
$$

From the above it follows that

$$
(15)\qquad \frac{\mu'_{[2]}}{2!} = (13) + (14).
$$

When $F = G$, this reduces to the expression known.

(8) *Mann and Whitney's T-statistic.* In this case the expression for the second factorial moment reduces to the simple form

$$
(16)\quad
\begin{aligned}
\mu'_{[2]} = 2n_1 n_2 \int_{-\infty}^{+\infty} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} & [(n_1 - 1) f(x_1) f(x_2) g(x_3) + (n_2 - 1) f(x_1) g(x_2) g(x_3)] \\
& \cdot dx_1 \, dx_2 \, dx_3 + n_1^{(2)} n_2^{(2)} \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{x_2} f(x_1) g(x_2) \, dx_1 \, dx_2 \right]^2,
\end{aligned}
$$

where $f(x)$ and $g(x)$ are the density functions for F and G.

(9) *AB joins between successive observations for a simple Markoff chain.* Let the matrix of probabilities for a simple Markoff chain be

$$
\begin{bmatrix} p_1 & p_2 \\ q_1 & q_2 \end{bmatrix}.
$$

Taking the probability that the first observation is A or B as P and Q respectively, the probabilities $P_r(A)$ and $Q_r(B)$ that the rth observation is A or B are given by

$$
(17)\quad
\begin{aligned}
P_r(A) &= \frac{p_2}{1 - \delta} + \frac{Pq_1 - Qp_2}{1 - \delta} \delta^{r-1}, \\
Q_r(B) &= 1 - P_r(A),
\end{aligned}
$$

where

$$\delta = p_1 - p_2 \quad \text{and} \quad p_1 > p_2 .$$

When the first observation is B, the conditional probability for the rth observation to be A reduces to

$$P_r(A \mid 1B) = \frac{p_2}{1-\delta}(1 - \delta^{r-1}). \tag{18}$$

This is the same as given by Bartlett:

In the case of the Markoff chain, unlike the previous cases discussed earlier, the probability of an AB join depends on the position of A in the sequence, and the expectation for two AB joins is given by

$$\begin{aligned} &q_1^2[P_1(A)\{P_3(A \mid 2B) + P_4(A \mid 2B) + P_5(A \mid 2B) + \cdots + P_{n-1}(A \mid 2B)\} \\ &+ P_2(A)\{P_4(A \mid 3B) + P_5(A \mid 3B) + P_6(A \mid 3B) + \cdots + P_{n-1}(A \mid 3B)\} \\ &+ P_3(A)\{P_5(A \mid 4B) + P_6(A \mid 4B) + \cdots + P_{n-1}(A \mid 4B)\} \\ &\qquad \cdots \\ &+ P_{n-3}(A)\{P_{n-1}(A \mid n-2, B)\}], \end{aligned} \tag{19}$$

where $P_r(A \mid B)$ is the conditional probability that the rth observation is A, given that the sth observation is B when $r > s$. Summing up the above series after substituting for P's from (16) and (17), we get

$$\begin{aligned} \frac{\mu'_{[2]}}{2!} = \Bigg[&\frac{(n-2)(n-3)\alpha}{2} - \frac{\alpha\delta}{1-\delta}\left\{(n-3) - \frac{\delta(1-\delta^{n-3})}{1-\delta}\right\} \\ &- \frac{\beta\delta}{1-\delta}\left\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\delta^{n-3}\right\} \\ &+ \beta\left\{\frac{n-3}{1-\delta} - \frac{\delta(1-\delta^{n-3})}{(1-\delta)^2}\right\}\Bigg] \frac{p_2 q_1^2}{1-\delta}, \end{aligned} \tag{20}$$

where

$$\alpha = \frac{p_2}{1-\delta} \quad \text{and} \quad \beta = \frac{Pq_1 - Qp_2}{1-\delta}.$$

It may be added that the result given in this paper can be used for deriving the moments of many other distributions of similar kind.

REFERENCES

1. M. S. BARTLETT, *An Introduction to Stochastic Processes*, Cambridge University Press, London, 1955, p. 29.
2. W. J. DIXON, "A criterion for testing the hypothesis that two samples are from the same population", *Ann. Math. Stat.*, Vol. 11 (1940), p. 199.
3. P. V. KRISHNA IYER, "The theory of probability distributions of points on a line", *J. Indian. Soc. Agri. Stat.*, Vol. 1 (1948), p. 173.

4. P. V. Krishna Iyer, "The first and second moments of some probability distribution arising from points on a lattice and their applications", *Biometrika*, Vol. 36 (1949), p. 135.

5. H. B. Mann and D. R. Whitney, "On a test whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), p. 50.

6. A. M. Mood, "On the asymptotic efficiency of certain nonparametric two sample tests", *Ann. Math. Stat.*, Vol. 25 (1954), p. 514.

7. A. Wald and J. Wolfowitz, "On a test whether two samples are from the same population," *Ann. Math. Stat.*, Vol. 11 (1940), p. 147.

THE UNIQUENESS OF THE TRIANGULAR ASSOCIATION SCHEME

By W. S. Connor

National Bureau of Standards

1. Summary. Parameters for a class of partially balanced incomplete block designs with two associate classes are immediately implied by the triangular association scheme. This paper deals with the more difficult question of whether or not these parameters imply the triangular association scheme.

2. Introduction. A partially balanced incomplete block design with two associate classes [1] is said to be triangular [2], [3] if the number of treatments $v = n(n-1)/2$ and the association scheme is an array of n rows and n columns with the following properties:

(a) The positions in the principal diagonal are blank.

(b) The $n(n-1)/2$ positions above the principal diagonal are filled by the numbers $1, 2, \cdots, n(n-1)/2$ corresponding to the treatments.

(c) The positions below the principal diagonal are filled so that the array is symmetrical about the principal diagonal.

(d) For any treatment i the first associates are exactly those treatments which lie in the same row and the same column as i.

The following relations clearly hold:

(1) The number of first associates of any treatment is $n_1 = 2n - 4$.

(2) With respect to any two treatments θ_1 and θ_2 which are first associates, the number of treatments which are first associates of both θ_1 and θ_2 is

$$p_{11}^1(\theta_1, \theta_2) = n - 2.$$

(3) With respect to any two treatments θ_3 and θ_4 which are second associates, the number of treatments which are first associates of both θ_3 and θ_4 is $p_{11}^2(\theta_3, \theta_4) = 4$.

We wish to examine the converse, i.e., whether or not relations (1), (2) and (3) imply (a), (b), (c), and (d). We shall give a proof for $n \geq 9$ which shows that the converse is true. The cases with $n < 9$ will not be considered, although the author has found that it is true for several small values of n, and conjectures that it is true for the rest.

As background for this problem, it is interesting to recall what has been found for some other classes of partially balanced designs. In the analogous problem for the group divisible designs it is easy to show that the converse is true [4]. For the latin square designs the converse is true for a sufficiently large number of treatments, but is not always true, as has been shown by example [5].

The present problem is closely related to problems considered in [6] and [7].

Received July 8, 1957.

The arguments used here could be substituted for some of the arguments in those papers.

3. A characterization of the triangular association scheme. The proof will consist of showing that there exist sets of treatments which satisfy the following theorem.

THEOREM. *The triangular association scheme for $n(n-1)/2$ treatments exists if and only if there exist sets of treatments S_j, $j = 1, \cdots, n$, such that:*

(i) *Each S_j consists of $n - 1$ treatments.*

(ii) *Any treatment is in precisely two sets S_j.*

(iii) *Any two distinct sets S_i, S_j have exactly one treatment in common.*

Proof. Necessity follows from the observation that the n rows of treatments in the triangular association scheme are the n sets S_j.

Sufficiency follows from noting a correspondence between the rows and columns of the association scheme and the sets S_j. To display the correspondence, we denote the unique element common to sets S_i and S_j by $\alpha(i, j) = \alpha(j, i)$. Then the correspondence is as follows: We let set S_i correspond to the ith row and column, and put element $\alpha(i, j)$ in the ith row and jth column of the association scheme. Because $\alpha(i_1, j_1) = \alpha(i_2, j_2)$ implies that $i_1 = i_2$ and $j_1 = j_2$, the element $\alpha(i, j)$ occurs only in the ith row (column) and jth column (row). This fills up the association scheme as described in (a), (b) and (c). Further, if we let "belonging to the same set S_j" correspond to "being first associates", then (d) is satisfied.

4. The existence of sets S_j which satisfy the Theorem. In this section we shall show for $n \geqq 9$ that there exist sets S_j which satisfy the Theorem. The proof makes conspicuous use of the condition (3) that $p_{11}^2 = 4$. In fact, in constructing the proof, the author was attracted to the singular fact that this parameter does not depend on n.

Throughout the proof, we shall employ certain conventions. In citing a reason why something is or is not true, we often shall write "$p_{11}^1(\theta_1, \theta_2)$" or "$p_{11}^2(\theta_3, \theta_4)$," whereby we mean to refer to particular treatments θ_1, θ_2, θ_3, and θ_4. Also, we shall write "$(\theta_1, \theta_2) = 1$ (or 2)," meaning that treatments θ_1 and θ_2 are first (or second) associates.

In developing the proof, the author used a matrix in which the ith row and column correspond to the ith treatment, and the entry in the intersection of the ith row and jth column is 1 or 2, depending on whether treatments i and j are first or second associates. Though this matrix is not explicitly used below, it is implicit, and it is believed that the reader will find the use of this matrix helpful in following the proof.

We begin by proving a lemma which will be used repeatedly in the sequel.

LEMMA 1. *With respect to any two initial treatments θ_1 and θ_2 which are first associates, the $n - 3$, $(n \geqq 9)$ treatments which are first associates of θ_1 and second associates of θ_2 pairwise are first associates.*

Proof. For simplicity we shall replace θ_1 by 1 and θ_2 by 2. From (1) and (2) it follows that there are $n - 2$ treatments which are first associates of both treatments 1 and 2, and $n - 3$ treatments which are first associates of treatment 1 and second associates of treatment 2. We shall refer to the treatments of the first set as treatments $3, \cdots, n$; and to those of the second set as treatments $n + 1, \cdots, 2n - 3$. These sets will be denoted respectively by

$$T_1 = T_1(3, \cdots, n)$$

and $T_2 = T_2(n + 1, \cdots, 2n - 3)$.

We first show that any treatment α in T_2 cannot have more than one second associate in T_2. We observe that $p_{11}^2(2, \alpha) = 4$, of which one such treatment is treatment 1. Thus, treatment α has at most three first associates in T_1. Because $p_{11}^1(1, \alpha) = n - 2$, treatment α has at least $n - 5$ first associates in T_2, and hence at most one second associate in T_2.

We now shall show that even this one second associate is impossible. Consider any two treatments α and β in T_2, and assume that $(\alpha, \beta) = 2$. We have established that treatment 1 and the $n - 5$ treatments other than α and β in T_2 are first associates of both α and β. But for $n \geqq 9$ the condition that $p_{11}^2(\alpha, \beta) = 4$ is violated, which shows that $(\alpha, \beta) = 1$. This completes the proof of Lemma 1.

Our next lemma shows the existence of sets S_j which satisfy (i) and (ii) of the theorem.

LEMMA 2. *For $n \geqq 9$, any initial treatment θ is an element of exactly two sets of treatments S_1 and S_2 which are such that a set contains $n - 1$ treatments, the treatments in a set pairwise are first associates, and θ is the unique element common to S_1 and S_2.*

Proof. We begin by showing that Lemma 1 implies that there are $n - 4$ treatments in T_1 which pairwise are first associates. For this purpose, it is convenient to define sets $T_1' = T_1'(3, \cdots, n - 2)$ and

$$T_2' = T_2'(n + 2, \cdots, 2n - 3).$$

From Lemma 1 and the condition that $p_{11}^1(1, \alpha) = n - 2$ for every treatment α in T_2, it follows that every treatment in T_2 has two first associates and $n - 4$ second associates in T_1. Without essential loss of generality, let treatment $n + 1$ be a second associate of every treatment in T_1', and let $(n - 1, n + 1) = (n, n + 1) = 1$. Then by Lemma 1, letting $\theta_1 = 1$ and $\theta_2 = n + 1$, the treatments in T_1' pairwise are first associates.

We still have to determine how treatments $n - 1$ and n intersect the treatments in T_1', T_2' and each other. We shall show that $(n - 1, n) = 2$ and either we have Case 1: $(n - 1, \alpha) = 1$, $(n, \alpha) = 2$ for all treatments α in T_1' and $(n - 1, \beta) = 2$, $(n, \beta) = 1$ for all treatments β in T_2'; or we have Case 2: $(n - 1, \alpha) = 2$, $(n, \alpha) = 1$ for all α in T_1' and $(n - 1, \beta) = 1$, $(n, \beta) = 2$ for all β in T_2'.

Suppose that treatment $n - 1$ is a second associate of some treatment in T_2', say treatment β. We shall show that we have Case 1. From Lemma 1 and $p_{11}^1(1, \beta)$ it follows that treatment β has two first associates and $n - 4$ second associates among the treatments of T_1' and treatments $n - 1$ and n. Therefore, treatment β has at least $n - 6$ second associates in T_1'. Without essential loss of generality, let these be treatments $3, \cdots, n - 4$. Applying Lemma 1 with $\theta_1 = 1$ and $\theta_2 = \beta$, it follows that these treatments are first associates of treatment $n - 1$.

Suppose that $(n - 3, n - 1) = 2$. Then it would be necessary that $p_{11}^2(n - 3, n - 1) = 4$. However, treatments $1, \cdots, n - 4$ are first associates of both treatments $n - 3$ and $n - 1$, violating $p_{11}^2(n - 3, n - 1) = 4$ for $n \geqq 9$. Similarly, treatment $n - 2$ cannot be a second associate of treatment $n - 1$.

We have shown that if treatment $n - 1$ has a second associate in T_2', then it is a first associate of every treatment in T_1'. Further, the treatments in T_1' and treatments 2 and $n + 1$ satisfy the condition that $p_{11}^1(1, n - 1) = n - 2$, implying that treatment $n - 1$ is a second associate of treatment n and the treatments in T_2'.

By applying Lemma 1 with $\theta_1 = 1$ and $\theta_2 = n - 1$, it follows that $(n, \beta) = 1$ for all β in T_2'. Because treatment 2 and the treatments in T_2 satisfy $p_{11}^1(1, n)$, it follows that $(n, \alpha) = 2$ for all α in T_1'. This demonstrates Case 1.

If $(n - 1, \beta) \neq 2$ for any β in T_2', then $(n - 1, \beta) = 1$ for all β in T_2'. But treatment 2 and the treatments in T_2 satisfy $p_{11}^1(1, n - 1)$, and it follows that $(n - 1, \alpha) = 2$ for all α in T_1' and that $(n - 1, n) = 2$. We now apply Lemma 1 with $\theta_1 = 1$ and $\theta_2 = n - 1$ to show that $(n, \alpha) = 1$ for all α in T_1'. Because treatments $2, \cdots, n - 2, n + 1$ satisfy $p_{11}^1(1, n) = n - 2$, it follows that $(n, \beta) = 2$ for all β in T_2'. This establishes Case 2.

We now observe a set S_1 which contains treatments 1, 2, the treatments in T_1' and either treatment $(n - 1)$ or n. Also, a set S_2 which contains treatments 1, the treatments in T_2, and the one of treatments $n - 1$ and n which is not in S_1. These sets are such that their elements pairwise are first associates. They are the sets of Lemma 2.

To show that there are no other such sets, we shall consider the way in which the treatments in T_1' are associated with the treatments in T_2'. Consider any treatment α in T_1', and the condition $p_{11}^1(1, \alpha) = n - 2$. Treatment 2, the remaining $(n - 5)$ treatments in T_1', and either treatment $n - 1$ or n are $n - 3$ treatments which satisfy this condition. Hence there is exactly one more such treatment in T_2'. Similarly, any treatment β in T_2' has exactly one first associate in T_1'. It follows that no other set of $n - 1$ treatments exists such that its treatments pairwise are first associates. This completes the proof of Lemma 2.

The sets found in Lemma 2 obey (i) and (ii) of the Theorem. To find the number s of sets and to prove (iii), we observe that each of s sets contains $n - 1$ elements, so that there are $s(n - 1)$ (not necessarily distinct) elements in the s sets. But every treatment occurs in exactly two sets, so that $s(n - 1) = 2v = n(n - 1)$ or $s = n$. Thus the number of pairs of sets is $n(n - 1)/2 = v$, and because every treatment occurs in exactly two sets, we have (iii).

REFERENCES

[1] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhya*, Vol. 4 (1939), pp. 337–372.

[2] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–190.

[3] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, *Tables of Partially Balanced Designs with Two Associate Classes*, Institute of Statistics of The Consolidated University of North Carolina Reprint Series No. 50, 1954, p. 43.

[4] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[5] Dale Marsh Mesner, *An Investigation of Certain Combinatorial Properties of Partially Balanced Incomplete Block Experimental Designs and Association Schemes, with a Detailed Study of Designs of Latin Square and Related Types*, unpublished thesis, Michigan State University, 1956.

[6] W. S. Connor, Jr., "On the structure of balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 57–71.

[7] Marshall Hall, Jr., and W. S. Connor, "An embedding theorem for balanced incomplete block designs," *Canadian Journal of Mathematics*, Vol. 6 (1953), pp. 35–41.

THE LIMITING DISTRIBUTION OF BROWNIAN MOTION IN A BOUNDED REGION WITH INSTANTANEOUS RETURN[1]

By B. Sherman

Westinghouse Research Laboratories

1. Summary. A point executes Brownian motion in a bounded, connected, and open three dimensional region D. When it reaches the boundary Γ, at point α, it is instantaneously returned to D according to probability measure $\mu(\alpha)$ (we write $\mu(\alpha, A)$ for the measure of set A), and the Brownian motion is resumed. This is a Markov process and, subject to certain regularity conditions on Γ and $\mu(\alpha)$, we derive the limiting distribution of the process. Processes of this sort have been considered by Feller [1]; he has obtained the transition probabilities of such processes. He is concerned more generally with Markov processes with continuous sample functions on a linear interval; the return may be instantaneous or after a random period of time.

Let $p^0(t, \xi, A)$ be the probability that the point is in set A of D at time t when it is initially at point ξ of D, with the additional restriction that no boundary contacts have been made. It is known that

$$p^0(t, \xi, A) = \int_A u(t, \xi, x)\, dx, \tag{1}$$

where dx is the volume element about x and u is the solution of the equation

$$\tfrac{1}{2}\Delta u = u_t,$$

subject to the conditions

$$u(t, \xi, \alpha) = 0, \qquad \alpha \varepsilon \Gamma, \qquad \lim_{t\to 0} \int_C u(t, \xi, x)\, dx = 1,$$

where C is any sphere of non-zero radius with center ξ which is entirely within D. We may write explicitly

$$u(t,\xi,x) = \sum_{k=1}^{\infty} v_k(\xi) v_k(x) e^{-\lambda_k t},$$

where λ_k is the kth eigenvalue and $v_k(x)$ the corresponding eigenfunction of the equation $\Delta u + 2\lambda u = 0$ subject to the boundary condition $u = 0$ on Γ. If $K(\xi, x)$ is the Green's function of $\Delta u = 0$ in D, then[2] ([2], and [3], page 273)

Received February 9, 1956; revised June 27, 1957.

[1] Part of this paper was prepared under the sponsorship of the Office of Naval Research and the Office of Ordnance Research, U.S. Army, at the University of California, Berkeley and Los Angeles. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] I am indebted to Professors M. Kac and R. J. Duffin for some helpful discussions on the relation of the fundamental solution of the heat equation to the Green's function of the Laplace equation.

$$K(\xi, x) = \frac{1}{2}\int_0^\infty u(t, \xi, x)\, dt. \tag{2}$$

Let $\phi(t, \xi, \alpha)\, dt\, d\alpha$ be the probability the point is absorbed at surface element $d\alpha$ of Γ between t and $t + dt$ when it is initially at point ξ of D. Then ϕ is half the interior normal derivative of u at point α of Γ([3], page 273). When the point is initially at ξ the probability of ultimate absorption in set S of Γ is given by

$$\pi(\xi, S) = \int_S \int_0^\infty \phi(t, \xi, \alpha)\, dt\, d\alpha. \tag{3}$$

We may define a discrete parameter Markov process with Γ as state space by taking as transition probability

$$\pi(\alpha, S) = \int_D \pi(\xi, S)\mu(\alpha, d\xi). \tag{4}$$

This Markov process has a limiting distribution π which satisfies the equation

$$\pi(S) = \int_\Gamma \pi(\alpha, S)\pi(d\alpha). \tag{5}$$

We define a measure of sets of D by

$$\lambda(A) = \int_\Gamma \mu(\alpha, A)\pi(d\alpha).$$

We may now write the density function for the limiting distribution. If $M(\xi)$ is the mean time of reaching the boundary when the point is initially at ξ,

$$M(\xi) = \int_\Gamma \int_0^\infty t\phi(t, \xi, \alpha)\, dt\, d\alpha$$

then the density function of the limiting distribution is

$$\frac{2\int_D K(\xi, x)\lambda(d\xi)}{\int_D M(\xi)\lambda(d\xi)}. \tag{6}$$

If we are given a probability measure λ in D and the return is always according to λ, then it is clear that the limiting density of this process is also given by (6). If λ concentrates at a single point ξ we may drop the integrals in (6), and in particular we get

$$M(\xi) = 2\int_D K(\xi, x)\, dx.$$

We note that (6) is essentially the steady distribution of temperature in the following problem: D is a homogeneous heat conducting body whose boundary is kept at temperature 0 and in which there is a constant source of heat distributed according to λ.

Regarding the regularity conditions, we shall assume that Γ is made up of finitely many surfaces, each with a continuously turning tangent plane and that D has a Green's function ([4], page 262). We will assume there is a closed set B in D such that

$$\inf_{\alpha \epsilon \Gamma} \mu(\alpha, B) = \gamma > 0.$$

2. Origin of the problem. This problem had its origins in the ecological research of Professor Thomas Park of the University of Chicago. He has been investigating problems of population stability and inter-species competition of flour beetles. It was discovered, on statistical investigations suggested in part by Jerzy Neyman, that the distribution of the beetles in the container of flour was not uniform, with the density increasing toward the boundaries of the container. The problem arose as to whether the nonuniformity might be simply a consequence of the random motion of the beetles or whether it ought to be attributed to some inhomogeneity such as a temperature gradient in the flour. To check the plausibility of the idea that the nonuniformity might arise from random motion alone, we have set up a model which may have some relevance to the actual situation. The region D represents the volume of flour. We assume the independence of the motions of the beetles so that we may confine ourselves to the random motion of a single point. This is a reasonable assumption if the density of the beetles is low. For the random motion we take Brownian motion; this is appropriate if we want path continuity and spatial homogeneity. Finally, we must introduce some mechanism of return from the boundary; we use the device of instantaneous return. If the return distribution is concentrated near the point of contact on the boundary then the device has some semblance of plausibility. More precisely we may suppose $\mu(\alpha, A(\alpha)) = 1$, where $A(\alpha)$ is that set of points of D whose distance from α is less than or equal to δ, a small positive number. Then if E is the subset of points of D whose distance from Γ is in excess of δ we have $\lambda(E) = 0$. If we are prepared to accept the density of distribution in E, as given by (6), as a theoretical model for what is observed then we are faced with a contradiction. For the density is a harmonic function in E, by virtue of $\lambda(E) = 0$. Because of the minimum-maximum properties of such functions we cannot have increasing density from the central parts of E outward to the boundary of E since that would entail a minimum at an interior point of E.

3. Derivation of the limiting distribution. We sketch a proof that we have defined a process by the instantaneous return mechanism. This is equivalent to proving that finitely many contacts occur in a finite time with probability 1. By the assumption on μ it will happen infinitely often with probability 1 that the point is returned to B. Let $T(x)$ be the time to reach Γ, starting at x. If δ is positive Prob $(T(x) > \delta)$ is a continuous function of x which achieves a positive minimum on B. Thus of the times the point is returned to B it will happen infinitely often with probability 1 that the time to reach Γ is in excess of δ. This implies that infinitely many contacts in a finite time has probability 0.

Let $p(t, x, A)$ be the transition probability of the process, i.e., the probability the point is in A at time t when it is initially at x. Then we prove the limiting distribution p exists and

$$p(t, x, A) = p(A) + f(t, x, A), \tag{7}$$

where

$$|f(t, x, A)| < ae^{-kt}. \tag{8}$$

Here a and k are positive and independent of x and A. To simplify the notation we will make the following convention: if $f(x)$ is a function on D then we define a corresponding function $f(\alpha)$ on Γ by taking the integral, over D, of f with respect to the measure $\mu(\alpha)$. With this convention we may replace x by α in (7) and (8). We note that both $f(t, x, A)$ and $f(t, \alpha, A)$ are integrable with respect to t from 0 to ∞.

Proceeding with the proof we use the fact that $u(t, \xi, x)$ is strictly positive for all ξ and x in D and for positive t. Then the minimum $v(x, \delta, t)$, achieved by $u(T, \xi, x)$ subject to $\xi \varepsilon B$ and $t - \delta \leqq T \leqq t$, is also strictly positive, and it follows directly that for $t - \delta \leqq T \leqq t$,

$$p^0(T, \alpha, A) \geqq \gamma \int_A v(x, \delta, t)\, dx.$$

It is clear that

$$h(\delta) = \inf_{x \varepsilon D} \int_\Gamma \int_0^\delta \phi(t, x, \alpha)\, dt\, d\alpha$$

satisfies $0 < h(\delta) < 1$ for all $\delta > 0$. Let $p^1(t, \xi, A)$ be the probability the point, initially at ξ, is in A at time t having made exactly one boundary contact. Then if $t > \delta$,

$$\begin{aligned} p^1(t, \xi, A) &= \int_\Gamma \int_0^t \phi(\tau, \xi, \alpha) p^0(t - \tau, \alpha, A)\, d\tau\, d\alpha \\ &\geqq \int_\Gamma \int_0^\delta \phi(\tau, \xi, \alpha) p^0(t - \tau, \alpha, A)\, d\tau\, d\alpha \\ &\geqq \int_\Gamma \int_0^\delta \phi(\tau, \xi, \alpha)\, d\tau\, d\alpha \cdot \gamma \int_A v(x, \delta, t)\, dx \\ &\geqq h(\delta) \gamma \int_A v(x, \delta, t)\, dx. \end{aligned}$$

We follow now the proof of a similar theorem given by Doob ([5], page 197). If $m(t, A)$ and $M(t, A)$ are respectively the infimum and supremum of $p(t, \xi, A)$ as ξ varies over D, then $M(t, A) \geqq m(t, A)$ and by the Chapman-Kolmogorov equation it can be seen that $M(t, A)$ is non-increasing and $m(t, A)$ is non-decreasing. For fixed t_0, ξ_0, x_0 define the set function

$$\psi(A) = p(t_0, \xi_0, A) - p(t_0, x_0, A).$$

There is a set A^+ on which ψ is maximum, such that $\psi(A) \geqq 0$ for any subset A of A^+, and such that $\psi(A) \leqq 0$ for any subset A of $A^- = D - A^+$. We have, assuming δ such that $0 < \delta < t_0$,

$$\begin{aligned}\psi(A^+) &= 1 - p(t_0, \xi_0, A^-) - p(t_0, x_0, A^+)\\ &\leqq 1 - p^1(t_0, \xi_0, A^-) - p^1(t_0, x_0, A^+)\\ &\leqq 1 - h(\delta)\gamma \int_D v(x, \delta, t_0)\, dx = c < 1.\end{aligned}$$

Following now a line of argument analogous to Doob's we have

$$M(t, A) - m(t, A) \leqq c^{(t/t_0)-1},$$

from which it follows that $M(t, A)$ and $m(t, A)$ have a common limit $p(A)$ and that

$$|p(t, x, A) - p(A)| \leqq M(t, A) - m(t, A) \leqq c^{(t/t_0)-1}.$$

Thus (7) and (8) are established, with $a = c^{-1}$ and $k = 1/t_0 \log c^{-1}$.

Before deriving (6) we have to establish the existence of the limiting distribution π of the boundary process. To this end we prove the lemma

$$\zeta = \sup_{S \epsilon \Gamma} (\max_{x \epsilon B} \pi(x, S) - \min_{x \epsilon B} \pi(x, S)) < 1. \tag{9}$$

We note that $\pi(x, S)$ is, for fixed S, a harmonic function of x([3], page 273). If (9) is not true there will be a sequence of sets S_k such that

$$\max_{x \epsilon B} \pi(x, S_k) \to 1, \qquad \min_{x \epsilon B} \pi(x, S_k) \to 0. \tag{10}$$

Since $\pi(x, S_k)$ is a sequence of harmonic functions with $0 \leqq \pi(x, S_k) \leqq 1$, we may extract a subsequence which converge to a harmonic function $f(x)$ uniformly on any compact subset of D([4], page 249). Without change of notation we suppose this done. However (10) implies that $f(x)$ achieves the values 1 and 0 on B, which contradicts the fact that $f(x)$ is harmonic in D and $0 \leqq f(x) \leqq 1$.

To prove the existence of the limiting distribution of the boundary process we again follow the lines of Doob's proof. For fixed α and β we define the set function

$$\psi(S) = \pi(\alpha, S) - \pi(\beta, S).$$

Associated with ψ are the sets S^+ and S^-, and we have

$$\begin{aligned}\psi(S^+) &= 1 - (\pi(\alpha, S^-) + \pi(\beta, S^+))\\ &\leqq 1 - \left(\int_B \pi(x, S^-)\mu(\alpha, dx) + \int_B \pi(x, S^+)\mu(\beta, dx)\right)\\ &\leqq 1 - \gamma\,(\min_{x \epsilon B} \pi(x, S^-) + \min_{x \epsilon B} \pi(x, S^+))\\ &= 1 - \gamma(1 - \max_{x \epsilon B} \pi(x, S^+) + \min_{x \epsilon B} \pi(x, S^+))\\ &\leqq 1 - \gamma(1 - \zeta).\end{aligned}$$

If now we introduce $m^{(n)}(S)$ and $M^{(n)}(S)$, the infimum and supremum of the n step transition probability $\pi^{(n)}(\alpha, S)$ as α ranges over Γ, then following Doob's proof

$$M^{(n)}(S) - m^{(n)}(S) \leqq (1 - \gamma(1 - \zeta))^n,$$

from which it follows that the limiting distribution exists and satisfies (5).

We are now in position to derive (6). We have

$$p(t, \xi, A) = p^0(t, \xi, A) + \int_\Gamma \int_0^t \phi(\tau, \xi, \alpha) p(t - \tau, \alpha, A) \, d\tau \, d\alpha.$$

Introducing (7) and integrating with respect to t we get, after some reductions,

$$\begin{aligned}\int_0^T p^0(t, \xi, A) \, dt = p(A)\Big[T\Big(1 - \int_0^T \int_\Gamma \phi(\tau, \xi, \alpha) \, d\alpha \, d\tau\Big)& \\ + \int_0^T \int_\Gamma \tau\phi(\tau, \xi, \alpha) \, d\alpha \, d\tau\Big] + I(T, \xi, A),&\end{aligned} \tag{11}$$

where

$$I(T, \xi, A) = \int_0^T f(t, \xi, A) \, dt - \int_0^T \int_\Gamma \int_0^t \phi(\tau, \xi, \alpha) f(t - \tau, \alpha, A) \, d\tau \, d\alpha \, dt. \tag{12}$$

The second term in the bracket on the right of (11) tends to $M(\xi)$ as $T \to \infty$, and we show that the first term tends to 0. This term can be written

$$T \text{ Prob } (x(t) \in D, 0 < t \leqq T \mid x(0) = \xi). \tag{13}$$

Let the coordinates of point x be x_1, x_2, x_3 and suppose D is contained between the planes $x_1 = a$ and $x_1 = -a$. Then (13) tends to 0 if the expression

$$T \text{ Prob } (-a < x_1(t) < a, 0 < t \leqq T \mid x(0) = \xi) \tag{14}$$

tends to 0. We may write (14) explicitly

$$T \sum_{n=1}^{\infty} \frac{4}{(2n+1)\pi} \sin \frac{(2n+1)\pi}{2a} (\xi + a) \exp\left(-\frac{(2n+1)^2\pi^2 T}{8a^2}\right),$$

which is less than

$$T \sum_{n=1}^{\infty} \frac{4}{(2n+1)\pi} \exp\left(-\frac{(2n+1)^2\pi^2 T}{8a^2}\right), \tag{15}$$

and it is easily proved that (15) tends to 0. Letting $T \to \infty$ in (11) we get

$$\int_0^\infty p^0(t, \xi, A) \, dt = p(A)M(\xi) + I(\infty, \xi, A). \tag{16}$$

Referring to (12) and (3) we may write, on introducing the variables $\tau' = \tau$ and $t' = t - \tau$,

$$I(\infty, \xi, A) = \int_0^\infty f(t, \xi, A)\, dt - \int_\Gamma \left(\int_0^\infty \phi(\tau', \xi, \alpha)\, d\tau'\right)\left(\int_0^\infty f(t', \alpha, A)\, dt'\right) d\alpha$$
$$= \int_0^\infty f(t, \xi, A)\, dt - \int_\Gamma \left(\int_0^\infty f(t, \alpha, A)\, dt\right) \pi(\xi, d\alpha).$$

The integration of the right side with respect to measure λ is equivalent to consecutive integrations with respect to $\mu(\beta)$ and π. The first integration gives, using (4),

$$\int_0^\infty f(t, \beta, A)\, dt - \int_\Gamma \left(\int_0^\infty f(t, \alpha, A)\, dt\right) \pi(\beta, d\alpha);$$

and the second, using (5), gives the value 0. Thus integrating on both sides of (16) with respect to measure λ we get

$$\int_D \int_0^\infty p^0(t, \xi, A)\, dt\, \lambda(d\xi) = p(A) \int_D M(\xi)\lambda(d\xi).$$

This equation, together with (1) and (2), implies that (6) is the density function of the limiting distribution p.

REFERENCES

[1] W. Feller, "Diffusion processes in one dimension," *Trans. Amer. Math. Soc.*, Vol. 77 (1954), pp. 1–31.

[2] G. A. Hunt, "On positive Green's functions," *Proc. Nat. Acad. Sci.*, Vol. 40 (1954), pp. 816–818.

[3] P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris, 1948.

[4] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, Vol. 2, Julius Springer, Berlin, 1937.

[5] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1953.

ON THE DISTRIBUTION OF A STATISTIC BASED ON ORDERED UNIFORM CHANCE VARIABLES

By Shanti S. Gupta[1] and Milton Sobel

Bell Telephone Laboratories

1. Summary. The exact distribution of a statistic based on the r smallest of n independent observations from a unit uniform distribution is derived. In life-testing terminology, this statistic includes as special cases (i) the sum of the r earliest failure times, (ii) the total observed life up to the rth failure, and (iii) the sum of all n failure times. The density, cumulative distribution function (c.d.f.) and first four moments of the general statistic are summarized in Sec. 2. Section 3 gives the derivation of the density and c.d.f. The moments are obtained from the moment generating function in Sec. 4. Asymptotic normality under certain conditions is proved in Sec. 5 and illustrations of the rapidity of approach to normality are given in Sec. 6.

2. Introduction and statement of results. We shall consider the statistic

$$T_{r,m}^{(n)} = t_1 + t_2 + \cdots + t_r + (m - r)t_r, \tag{2.1}$$

where $t_i = t_i^{(n)}$ is the ith smallest of n independent observations and m is greater than $r - 1$ but is not necessarily an integer. For $m = n$ this statistic can be interpreted as the total observed life in a life-testing experiment without replacement. When the underlying distribution of the unordered t's is exponential, i.e., $f(t) = (1/\theta)e^{-t/\theta}$, then it is known [3] that $2T_{r,n}^{(n)}/\theta$ is distributed as chi-square χ_{2r}^2 with $2r$ degrees of freedom.

Before stating further results let us introduce for $0 \leqq t \leqq m$ and non-negative integers p, q, n

$$A_{p,m}^{(q,n)}(t) = \frac{n}{p!}\left\{\binom{p}{0}\frac{t^{n-1}}{m^{q-p}} - \binom{p}{1}\frac{(t-1)^{n-1}}{(m-1)^{q-p}} + \binom{p}{2}\frac{(t-2)^{n-1}}{(m-2)^{q-p}} - \cdots\right\}, \tag{2.2}$$

where $m > p$, $n \geqq 1$ and the summation is continued as long as the arguments $t, t - 1, t - 2, \cdots$ are positive. It is understood that the binomial coefficient $\binom{p}{j} = 0$ for $p < j$ so that there are at most $(p + 1)$ terms in the above summation.

It is clear from (2.1) that $T_{n,n}^{(n)}$ is the sum of all the n observations. When the underlying distribution is unit uniform, then the density of $T_{n,n}^{(n)}$ is given on p. 246 of [2] by

$$f_{n,n}^{(n)}(t) = \frac{1}{(n-1)!}\left\{\binom{n}{0}t^{n-1} - \binom{n}{1}(t-1)^{n-1} + \cdots\right\} = A_{n,m}^{(n,n)}(t). \tag{2.3}$$

Received March 5, 1957, revised May 16, 1957.

[1] Now at the University of Alberta.

(We have removed the superscripts and subscripts from the chance variables and put them on f and on F below which are the symbols for the density and c.d.f., respectively.)

Using the symmetry of the above density about $t = n/2$, we can replace t by $n - t$ in (2.3) obtaining

$$
\begin{aligned}
f_{n,n}^{(n)}(t) &= \frac{n}{(n-1)!}\left\{\binom{n-1}{0}\frac{(n-t)^{n-1}}{n}\right. \\
&\quad \left. - \binom{n-1}{1}\frac{(n-1-t)^{n-1}}{(n-1)} + \cdots\right\} = A_{n-1,n}^{(n,n)}(n-t),
\end{aligned} \tag{2.4}
$$

where $0 \leqq t \leqq n$. The form (2.4) is more comparable with the results derived here. It is shown below that the density and c.d.f. of $T_{r,m}^{(n)}$ are given by the comparable results

$$f_{r,m}^{(n)}(t) = A_{r-1,m}^{(n,n)}(m - t) \tag{2.5}$$

and

$$F_{r,m}^{(n)}(t) = 1 - \frac{1}{(n+1)} A_{r-1,m}^{(n,n+1)}(m - t), \tag{2.6}$$

from which we get as special cases the densities and c.d.f.'s of (i) $T_{r,r}^{(n)}$, (ii) $T_{r,n}^{(n)}$ and (iii) $T_{n,n}^{(n)}$.

Barton and David [1] have derived another equivalent formula for the density $f_{r,r}^{(n)}(t)$, i.e., in the special case (i). Their result, with two typographical corrections taken into account, is

$$f_{r,r}^{(n)}(t) = \frac{n}{r!}\sum_{i=1}^{r}(-1)^{r-i} r^{r-n}\binom{r}{i}\left[\frac{i - t + |i - t|}{2}\right]^{n-1}. \tag{2.7}$$

The total life statistic arises as an optimum statistic under exponential distribution assumptions in [3]. In the present paper we give the distribution of this statistic when the exponential distribution assumption is replaced by the uniform distribution. Hence these results can be used to study the robustness of the tests based on the total life statistic. The results on asymptotic normality are also of interest in this connection since under the exponential assumption the distribution of $2T_{r,n}^{(n)}/\theta$ is that of χ_{2r}^2 which, for large r, also is close to that of a normal distribution. It is felt that the model of a uniform distribution from 0 to θ, $\theta > 0$ and unknown, and the results of this paper may prove to be useful in some life-testing problems.

3. Derivation of results. Let $u = t_1 + t_2 + \cdots + t_{r-1}$, $v = t_r$, $w = u/v$ and $y = T_{r,m}^{(n)} = u + (m - r + 1)v$, where $t_i = t_i^{(n)}$ is the ith smallest of n independent chance variables uniformly distributed from zero to one. The conditional distribution of w given v is exactly that of a sum of $r - 1$ independent uniform chance variables and is given by (2.3) with n replaced by $r - 1$. Hence the joint density of v and w is given by

$$
\begin{aligned}
g(w, v) &= \frac{n!}{(r-1)!(n-r)!} v^{r-1}(1-v)^{n-r} \frac{1}{(r-2)!} \\
&\quad \cdot \left\{ \binom{r-1}{0} w^{r-2} - \binom{r-1}{1} (w-1)^{r-2} + \cdots \right\},
\end{aligned} \tag{3.1}
$$

where $0 \leqq w \leqq r - 1$ and $0 < v \leqq 1$ and the joint density of u and v is given by

$$
\begin{aligned}
h(u, v) &= \frac{n!}{(r-2)!(r-1)!(n-r)!} \\
&\quad \cdot \left\{ \binom{r-1}{0} u^{r-2} - \binom{r-1}{1} (u-v)^{r-2} + \cdots \right\} (1-v)^{n-r},
\end{aligned} \tag{3.2}
$$

where $0 \leqq u \leqq (r-1)v$ and $v \leqq 1$.

If we now derive the density of y, then the full range of y from 0 to m is broken into r parts. For $0 \leqq y \leqq m - r + 1$, the density of y becomes

$$
\begin{aligned}
f_{r,m}^{(n)}(y) &= \frac{n!}{(r-2)!(r-1)!(n-r)!} \\
&\quad \cdot \int_{y/m}^{y/(m-r+1)} [y - (m-r+1)v]^{r-2}(1-v)^{n-r}\, dv \\
&\quad - \binom{r-1}{1} \int_{y/m}^{y/(m-r+2)} [y - (m-r+2)v]^{r-2}(1-v)^{n-r}\, dv \\
&\quad + \cdots + (-1)^{r-2} \binom{r-1}{r-2} \int_{y/m}^{y/(m-1)} [y - (m-1)v]^{r-2}(1-v)^{n-r}\, dv.
\end{aligned} \tag{3.3}
$$

Using the finite difference operators $\mathcal{E}$, Δ (with $\mathcal{E} = 1 + \Delta$),

$$
\begin{aligned}
f_{r,m}^{(n)}(y) &= \frac{r}{(r-2)!} \binom{n}{r} \sum_{\alpha=0}^{r-1} \binom{r-1}{\alpha} \\
&\quad \cdot (-\mathcal{E})^{\alpha} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y - (m-r+1+x)v]^{r-2}(1-v)^{n-r}\, dv \right\},
\end{aligned} \tag{3.4}
$$

where $\mathcal{E}$ operates on x and it is understood that x is then to be set equal to 0. Using the relation between $\mathcal{E}$ and Δ,

$$
\begin{aligned}
f_{r,m}^{(n)}(y) &= \frac{r}{(r-2)!} \binom{n}{r} \\
&\quad \cdot \left[(-\Delta)^{r-1} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y - (m-r+1+x)v]^{r-2}(1-v)^{n-r}\, dv \right\} \right]_{x=0}.
\end{aligned} \tag{3.5}
$$

If we now integrate by parts, the first term vanishes at the upper limit and also at the lower limit because of the operator Δ^{r-1}. After $r - 1$ such integrations we obtain

$$
f_{r,m}^{(n)}(y) = \frac{n}{(r-1)!} \left[\Delta^{r-1} \left\{ \frac{(m-r+1+x-y)^{n-1}}{(m-r+1+x)^{n-r+1}} \right\} \right]_{x=0}. \tag{3.6}
$$

Using $\Delta = \mathcal{E} - 1$ we obtain

$$f_{r,m}^{(n)}(y)$$

$$(3.7)\quad = \frac{n}{(r-1)!}\left\{\binom{r-1}{0}\frac{(m-y)^{n-1}}{m^{n-r+1}} - \binom{r-1}{1}\frac{m-1-y)^{n-1}}{(m-1)^{n-r+1}} + \cdots\right\}$$
$$= A_{r-1,m}^{(n,n)}(m-y),$$

where $0 \leqq y \leqq m - r + 1$ and $A_{p,m}^{(q,n)}$ is defined in (2.2).

We shall now show that the expression (3.7) gives the result for all $y(0 \leqq y \leqq m)$. For $m - r + i \leqq y \leqq m - r + 1 + i$ $(i = 1, 2, \cdots, r - 1)$ the only difference is that the first i upper limits of integration in (3.3) are all changed to unity. For the jth integral $(j = 1, 2, \cdots, i)$ we have to add to the complete set of r terms in (3.7) the quantity

$$(-1)^{j+1}\binom{r-1}{j-1}\frac{n!}{(r-2)!(r-1)!(n-r)!}$$

$$(3.8)\quad \cdot\int_{y/m-r+j)}^{1}[y-(m-r+j)v]^{r-2}(1-v)^{n-r}\,dv$$

$$= (-1)^{j+r-1}\frac{n}{(r-1)!}\binom{r-1}{j-1}\frac{(m-r+j-y)^{n-1}}{(m-r+j)^{n-r+1}}.$$

For each $j(1 \leqq j \leqq i)$ the quantity on the right in (3.8) cancels the jth term from the end of the complete expression with r terms in (3.7). Hence for

$$m - r + i \leqq y \leqq m - r + i + 1$$

the density is given by the first $r - i$ terms of (3.7) which are precisely those terms with positive arguments. This proves that the expression $A_{r-1,m}^{(n,n)}(m - y)$ of (3.7) gives the result for all $y(0 \leqq y \leqq m)$.

The c.d.f. $F_{r,m}^{(n)}(y)$ of y is easily obtained by integrating (3.7) between the limits 0 and y and is given by

$$(3.9)\quad F_{r,m}^{(n)}(y) = 1 - \frac{1}{(n+1)}A_{r-1,m}^{(n,n+1)}(m-y).$$

4. Moments of $y = T_{r,m}^{(n)}$. Using the expression for the density it can be shown that the moment generating function $M_\theta(y)$ of $y = T_{r,m}^{(n)}$ is given by

$$(4.1)\quad M_\theta(y) = \left[\frac{n}{(r-1)!}\sum_{\alpha=0}^{r-1}(-1)^\alpha\binom{r-1}{\alpha}\right.$$
$$\left.\cdot\mathcal{E}^\alpha\left\{\frac{1}{(m-x)^{n-r+1}}\int_0^{m-x}e^{\theta y}(m-x-y)^{n-1}\,dy\right\}\right]_{x=0}$$

$$(4.2)\quad = \frac{n!}{(r-1)!}\sum_{j=0}^{\infty}\frac{(-\theta)^j}{(j+n)!}[\Delta^{r-1}(x-m)^{j+r-1}]_{x=0}.$$

Thus we have for the jth moment

$$E(y^j) = \frac{(-1)^j j!n!}{(r-1)!(n+j)!}[\Delta^{r-1}\{(x-m)^{r-1+j}\}]_{x=0} \tag{4.3}$$

$$= \frac{(-1)^j j!n!}{(r-1)!(n+j)!}\sum_{\beta=0}^{j}(-1)^\beta m^\beta \binom{r-1+j}{\beta}[\Delta^{r-1}x^{r-1+j-\beta}]_{x=0}. \tag{4.4}$$

It can be shown that for $j \geqq 0$ and $r \geqq 1$

$$[\Delta^{r-1}x^{r+j}]_{x=0} = \sum_{\alpha=1}^{r-1}\frac{(r-1)!}{(\alpha-1)!}[\Delta^\alpha x^{\alpha+j}]_{x=0}. \tag{4.5}$$

The results for various values of j in (4.5) are known and are given, for example, in [4], p. 127. Using these we have from (4.4)

$$E(y) = \frac{r(2m-r+1)}{(2n+1)}, \tag{4.6}$$

$$E(y^2) = \frac{r(r+1)}{12(n+1)(n+2)}\left[12m^2 - 12m(r-1) + (r-1)(3r-2)\right], \tag{4.7}$$

$$\sigma^2(y) = \frac{r(n-r+1)(2m-r+1)^2}{4(n+1)^2(n+2)} + \frac{r(r+1)(r-1)}{12(n+1)(n+2)}, \tag{4.8}$$

$$\begin{aligned} E(y^3) &= \frac{r(r+1)(r+2)}{8(n+1)(n+2)(n+3)} \\ &\quad \cdot \left[8m^3 - 12m^2(r-1) + 2m(r-1)(3r-2) - r(r-1)^2\right], \end{aligned} \tag{4.9}$$

$$\begin{aligned} E(y^4) &= \frac{r(r+1)(r+2)(r+3)}{2(n+1)(n+2)(n+3)(n+4)} \\ &\quad \cdot \Big[2m^4 - 4m^3(r-1) + m^2(r-1)(3r-2) - m(r-1)^2 r \\ &\qquad + \frac{(r-1)(15r^3 - 15r^2 - 10r + 8)}{120}\Big]. \end{aligned} \tag{4.10}$$

Since the computation of cumulants leads to no simplification, they have not been given here; they can be obtained by the usual formulae. It should be mentioned that the above expressions for the moments can also be obtained directly by using the moments of the order statistics.

5. Asymptotic normality of $y = T_{r,m}^{(n)}$. We shall randomize the order of the chance variables $t_1, t_2, \cdots, t_{r-1}$ and thus define new unordered *equi-correlated* and identically distributed chance variables $u_1, u_2, \cdots, u_{r-1}$. Furthermore, if we consider the conditional joint distribution of the u_i given v $(=t_r)$, then we have *independent* chance variables which are uniformly distributed from 0 to v.

Let

$$y^* = \frac{y - E(y)}{\sigma(y)} \text{ and } y_v^* = \frac{y - E(y \mid v)}{\sigma(y \mid v)}, \tag{5.1}$$

where $y = T_{r,m}^{(n)} = u_1 + u_2 + \cdots + u_{r-1} + (m - r + 1)v$.
The characteristic function of y^* is given by

$$\begin{aligned} \varphi_y^*(t) &= \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\left(\frac{y - E(y)}{\sigma(y)}\right)\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\, dv \\ &= \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\, \frac{\sigma(y \mid v)}{\sigma(y)}\left[\frac{y - E(y \mid v)}{\sigma(y \mid v)}\right]\right. \\ &\quad \left. + it\, \frac{\sigma(v)}{\sigma(y)}\left[\frac{E(y \mid v) - E(y)}{\sigma(v)}\right]\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\, dv, \end{aligned} \tag{5.2}$$

where

$$E(y \mid v) = (r - 1)\frac{v}{2} + (m - r + 1)v = (2m - r + 1)\frac{v}{2}, \tag{5.3}$$

$$\sigma(y \mid v) = v\sqrt{\frac{r - 1}{12}}; \qquad \sigma(v) = \frac{1}{(n + 1)}\sqrt{\frac{r(n - r + 1)}{n + 2}}, \tag{5.4}$$

and

$$g(v) = \frac{n!}{(r - 1)!(n - r)!} v^{r-1}(1 - v)^{n-r}. \tag{5.5}$$

Letting

$$t' = t\,\frac{\sigma(y \mid v)}{\sigma(y)}, \qquad t'' = t\,\frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r - 1}{2}\right)\right], \tag{5.6}$$

and $x_i = u_i/v(i = 1, 2, \cdots, r - 1)$, we obtain

$$\varphi_y^*(t) = \int_0^1 \left[\int_0^1 e^{it'\sqrt{\frac{12}{r-1}}(x-\frac{1}{2})}\, dx\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\, dv \tag{5.7}$$

$$= \int_0^1 \left[\frac{\sin\left(t'\sqrt{\frac{3}{r - 1}}\right)}{t'\sqrt{\frac{3}{r - 1}}}\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\, dv. \tag{5.8}$$

Since for $r = \lambda n$ and $n \to \infty$ we have

$$E(v) = \frac{r}{n + 1} \to \lambda \quad \text{and} \quad \sigma(v) = \frac{1}{n + 1}\sqrt{\frac{r(n - r + 1)}{n + 2}} = O\left(\frac{1}{\sqrt{n}}\right), \tag{5.9}$$

then we shall write $v = \lambda + O(1/\sqrt{n})$ in the expression (5.6) for t' which is needed for the first part of the integrand in (5.8). For $m = \gamma n$ we obtain the two asymptotic relations

$$
(5.10)\qquad \begin{aligned}\frac{\sigma(y \mid v)}{\sigma(y)} &\cong \frac{v}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}} \\ &= \frac{\sqrt{\lambda^2}}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}} + O\left(\frac{1}{\sqrt{n}}\right)\end{aligned}
$$

and

$$
(5.11)\qquad \frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r-1}{2}\right)\right] \cong \frac{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2}}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}},
$$

so that if we denote the first term in the right hand members of (5.10) and (5.11) by a and b respectively, then $a^2 + b^2 = 1$. Taking the limit in (5.8) as $n \to \infty$ with $r = \lambda n$, $m = \gamma n$ and using the Lebesgue theorem, we can bring the limit operator under the integral sign. Then, using (5.10), we obtain

$$
(5.12)\qquad \varphi_y^*(t) \cong \int_0^1 \lim\left[1 - \frac{a^2t^2}{2(r-1)} + O\left(\frac{1}{n^{3/2}}\right)\right]^{r-1} \lim e^{it^*\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\,dv
$$

$$
(5.13)\qquad = e^{-\frac{a^2t^2}{2}} \int_0^1 \lim e^{it^*\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\,dv.
$$

Using the same Lebesgue theorem the limit operator can be taken outside the integral sign. Then, using a result on the asymptotic normality of quantiles given on page 369 in Cramér [2], we obtain

$$
(5.14)\qquad \varphi_y^*(t) \cong e^{-\frac{a^2t^2}{2}} e^{-\frac{b^2t^2}{2}} = e^{-\frac{t^2}{2}}
$$

since $a^2 + b^2 = 1$. This proves the asymptotic normality of y for $r = \lambda n$, $m = \gamma n$ (γ and λ fixed with $0 < \lambda \leqq 1$ and $\lambda \leqq \gamma < \infty$) and $n \to \infty$.

It should be noted that the above proof holds no matter how fast m tends to infinity. If $m/n \to \infty$ then $a = 0$ and $b = 1$ and (5.14) still holds.

6. Illustration of rapidity of approach to normality. To illustrate the rapidity of approach to normality of the statistics, we shall use the Edgeworth series expansion

$$
(6.1)\qquad \begin{aligned} F_{r,m}^{(n)}(x) = \{\Phi(x)\} &- \left\{\frac{1}{3!}\frac{\mu_3}{\sigma^3}\Phi^{(3)}(x)\right\} \\ &+ \left\{\frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right)\Phi^{(4)}(x) + \frac{10}{6!}\left(\frac{\mu_3}{\sigma^3}\right)^2\Phi^{(6)}(x)\right\} + \cdots, \end{aligned}
$$

where $\Phi(x)$ is the standard normal c.d.f., $\Phi^{(r)}(x)$ is its rth derivative and x denotes the standardized variate corresponding to t. We wish to compute one, two, and three terms of (6.1) as indicated by the braces for the two special cases of $T_{r,m}^{(n)}$; viz., (i) $m = r$ and (ii) $m = n$. These have been computed for $n = 10$, $r = 5$ and the results are compared in Table I below with the exact values computed from (2.6).

TABLE I

Comparison of exact probability $P[T_{r,m}^{(n)} \leq t]$ and Edgeworth approximations

Case	t	x	Approximations			Exact Probability
			1 term	2 terms	3 terms	
(i)	1.5	0.26656	.6051	.6340	.6318	.6327
$r = 5$	2.0	1.24393	.8932	.8851	.8849	.8839
$m = 5$	2.5	2.22131	.9868	.9761	.9780	.9769
$n = 10$	3.0	3.19868	.9993	.9975	.9969	.9973
(ii)	4.0	0.30754	.6208	.6312	.6261	.6259
$r = 5$	5.0	1.15329	.8756	.8736	.8681	.8671
$m = 10$	6.0	1.99902	.9772	.9723	.9741	.9739
$n = 10$	7.0	2.84475	.9978	.9963	.9976	.9979

7. Acknowledgment. The authors wish to thank Prof. J. W. Tukey for his helpful comments and suggestions.

REFERENCES

[1] D. E. Barton and F. N. David, "Sums of ordered intervals and distances," *Mathematika*, Vol. 2 (1955), pp. 150–159.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] B. Epstein and M. Sobel, "Some theorems relevant to life testing from an exponential distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 373–381.

[4] L. M. Milne-Thomson, *The Calculus of Finite Differences*, MacMillan, London, 1933.

DETERMINING SAMPLE SIZE FOR A SPECIFIED WIDTH CONFIDENCE INTERVAL

By Franklin A. Graybill

Oklahoma State University

1. Introduction. If an experimenter decides to use a confidence interval to locate a parameter, he is concerned with at least two things: (1) Does the interval contain the parameter? (2) How wide is the interval? In general the answer to these questions cannot be given with absolute certainty, but must be given with a probability statement. If we let α be the probability that the interval contains the parameter, and let β^2 be the probability that the width is less than d units, then the general procedure is to fix α in advance and compute β^2. The value of β^2 is in general a function of the positive integer n, the sample size by which the confidence interval is computed. (β^2 is also a function of α). In most confidence intervals, β^2 increases as n increases. For any particular situation β^2 may be too low to be useful, hence an experimenter may wish to increase β^2 by taking more observations (increasing n). The problem the experimenter then faces is the determination of n such that (A) the probability will be equal to α that the confidence interval contains the parameter, and (B) the probability will be equal to β^2 that the width of the confidence interval will be less than d units (where α, β^2, and d are specified).

To solve this problem will generally require two things: (1) The form of the frequency function from which the sample of size n is to be selected; (2) Some previous information on the unknown parameters in the frequency function.

This suggests that the sample be taken in two steps; the first sample will be used to determine the number of observations to be taken in the second sample so that (A) and (B) will be satisfied.

For a confidence interval on the mean of a normal population with unknown variance this problem has been solved by Stein [1] for $\beta^2 = 1$.

The purpose of this paper is to determine n, to satisfy (A) and (B) for distributions other than the normal.

2. Theory. Suppose X is the width of a confidence interval on a parameter μ with confidence coefficient α. Suppose further that it is desired that the probability be β^2 that X be less than d. The problem is to determine n, the number of observations, on which to base X. Since n depends on the random variables used in step one, n is a random variable.

We will prove the following (we will use the notation $P(A)$ for the probability that the event A occurs):

Theorem. *Let the chance variable X be the width of a confidence interval on a parameter μ based on a sample of size n. Suppose that X depends on n and on an unknown parameter θ (θ may be the parameter μ). Suppose also that there exists a*

Received April 29, 1957; revised October 8, 1957.

function of X, θ, and n, say $g(X; \theta, n)$, such that if $Y = g(X; \theta, n)$, then the distribution of Y does not depend on any unknown parameters except n. Let $f(n)$ be a function of n such that

$$(1) \qquad P[Y < f(n)] = \beta \quad \text{for any} \quad 0 < \beta < 1.$$

Let the solution of the equation $g(x; \theta, n), = f(n)$ for x be $x = h(\theta, n)$, and suppose the following are true for $x > 0$:

(2)
- (a) *$g(x; \theta, n)$ is monotonic increasing in x for every n and θ.*
- (b) *$h(\theta, n)$ is monotonic increasing for every n.*
- (c) *$h(\theta, n)$ is monotonic decreasing in n for every θ.*
- (d) *z is random variable which is available from step one of the procedure such that $P[t(z) > \theta] = \beta$ for $O < \beta < 1$, where $t(z)$ is a function of z which does not depend on any unknown parameters or on n.*

Let d and β be specified in advance. Then if n is such that the equation

$$(3) \qquad h[t(z), n] \leqq d$$

is satisfied ($t(z)$ is known) then the following inequality is true:

$$(4) \qquad P(X \leqq d) \geqq \beta^2.$$

PROOF. Substituting into Eq. (1) we get

$$(5) \qquad P[g(X; \theta, n) < f(n)] = \beta.$$

Solving for X and using 2(a) gives us

$$(6) \qquad P[X < h(\theta, n)] = \beta.$$

For any $\theta_1 \geqq \theta$ we can use 2(b) and obtain

$$(7) \qquad P[X < h(\theta_1, n) \mid \theta_1 \geqq \theta] \geqq P[X < h(\theta, n)] = \beta.$$

By considering the joint distribution of X and $t(z)$ we can write

$$(8) \quad P(X \leqq d) \geqq P[X \leqq d, t(z) > \theta] = P[X \leqq d \mid t(z) > \theta] \cdot P[t(z) > \theta].$$

If n is any integer satisfying

$$(9) \qquad h[t(z), n] \leqq d,$$

we can use (7), 2(c), and 2(d) in Eq. (8) and obtain

$$P(X \leqq d) \geqq \beta^2.$$

If the function in 2(b) is monotonic *decreasing*, then the theorem is also true but the inequality in 2(d) must be reversed. The theorem is also true if the function in 2(a) is monotonic *decreasing*. The conditions (2) may appear quite stringent; however, many of the functions in common use in statistics satisfy these conditions.

3. Illustrations

Example 1. Suppose we want an α confidence interval on the variance σ^2 of a normal population to be less than d units in length with probability of β^2.

We will define

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (v_i - \bar{v})^2 ,$$

where v_i is distributed normally with mean μ and variance σ^2. An α confidence interval on σ^2 is given by

$$P\left[\frac{(n-1)\, s_n^2}{\chi_1^2\,(n)} \leqq \sigma^2 \leqq \frac{(n-1)\, s_n^2}{\chi_2^2\,(n)}\right] = \alpha,$$

where $\chi_1^2(n)$ and $\chi_2^2(n)$ are such that

$$\int_0^{\chi_2^2(n)} W(\chi^2; n)\, d\chi^2 = \frac{1-\alpha}{2},$$

$$\int_{\chi_1^2(n)}^{\infty} W(\chi^2; n)\, d\chi^2 = \frac{1-\alpha}{2},$$

where $W(\chi^2; n)$ is a Chi-square frequency function with $n - 1$ degrees of freedom. The width of the interval is

$$X = (n-1)\, s_n^2\left[\frac{1}{\chi_2^2\,(n)} - \frac{1}{\chi_1^2\,(n)}\right].$$

If we let

$$\frac{1}{\chi_2^2\,(n)} - \frac{1}{\chi_1^2\,(n)} = C_n ,$$

we have $g(X; \sigma^2, n) = X/\sigma^2 C_n = Y$, and we see that Y is distributed as $W(\chi^2; n)$ and is independent of any unknown parameters except n.

Also $f(n)$ is given by $\int_0^{f(n)} W(\chi^2; n)\, d\chi^2 = \beta$, and $h(\theta, n) = \sigma^2 C_n f(n)$.

Suppose in step one of our procedure we observe $u_1, u_2, \cdots, u_m$ which is a random sample of size m from a normal population with variance σ^2. If we let

$$z = \sum_{i=1}^{m} (u_i - \bar{u})^2,$$

then since z/σ^2 is distributed as $W(\chi^2; m)$, it is clear that $P[(z/\sigma^2) > f_m] = \beta$, where f_m is such that

$$\int_{f_m}^{\infty} W(\chi^2; m)\, d\chi^2 = \beta .$$

Hence $t(z) = z/f_m$, and since all the conditions in (2) are satisfied, the sample size for the desired length of the confidence interval is the smallest integral value of n satisfying

$$\frac{f(n) \cdot C_n \cdot z}{f_m} \leqq d\,.$$

Example 2. Next, suppose it is desired to determine the sample size such that an α confidence interval on the mean of a normal population will have width less than d with probability β^2. Let $v_1, v_2, \cdots, v_n$ be a random sample of size n(to be determined) from a normal distribution with mean μ and variance $\sigma^2 = \theta^2$. If we let

$$s_n^2 = \frac{1}{n-1} \sum (v_i - \bar{v})^2,$$

then an α confidence interval on μ is

$$\bar{v} - \frac{t_0 s_n}{\sqrt{n}} \leqq \mu \leqq \bar{v} + \frac{t_0 s_n}{\sqrt{n}},$$

where t_0 is such that

$$\int_{t_0}^{\infty} U(t, n)\, dt = \frac{1-\alpha}{2},$$

where $U(t, n)$ is "Student's" distribution with $(n - 1)$ degrees of freedom. The length of the interval is

$$X = 2\,\frac{t_0\, s_n}{\sqrt{n}}\,.$$

If we let

$$Y = g(X; \theta, n) = \frac{(n-1)s_n^2}{\sigma^2} = \frac{n(n-1)X^2}{4t_0^2\, \sigma^2},$$

then Y is distributed as a Chi-square variate with $(n - 1)$ degrees of freedom, and is independent of any unknown parameters except n.

If $W(\chi^2; n)$ is a Chi-square frequency function with $(n - 1)$ degrees of freedom, then $f(n)$ is given by

$$\int_0^{f(n)} W(\chi^2; n)\, d\chi^2 = \beta,$$

and

$$X = h(\theta, n) = 2t_0\, \sigma\, \frac{\sqrt{f(n)}}{\sqrt{n(n-1)}}\,.$$

Suppose $u_1, u_2, \cdots, u_m$ is a random sample of size m from a normal population with variance σ^2 which is available from step one of our procedure.

If we let

$$z = \sum_{i=1}^{m} (u_i - \bar{u})^2,$$

then we have $P[(z/\sigma^2) > f_m] = \beta$, where f_m is such that $\int_{f_m}^{\infty} W(\chi^2; m)\, d\chi^2 = \beta$. Hence, $t(z) = (z/f_m)^{\frac{1}{2}}$, and since all the conditions in (2) are satisfied, the sample size for step two is the smallest integral value of n satisfying

$$\frac{2t_0\sqrt{z}}{\sqrt{f_m}} \cdot \frac{\sqrt{f(n)}}{\sqrt{n(n-1)}} \leqq d.$$

It is interesting to compare the method in this paper with the method presented by Stein [1] for setting a confidence interval on the mean of a normal population with unknown variance.

The procedure presented by Stein is to select a two step sample. Suppose the sample in the first step is $u_1, u_2, \cdots, u_m$ and is taken from a normal population with mean μ and variance σ^2. An α confidence interval on μ is

$$\bar{u} - \frac{t_m s}{\sqrt{m}} \leqq \mu \leqq \bar{u} + \frac{t_m s}{\sqrt{m}},$$

where $s^2 = 1/(m-1) \sum (u_i - \bar{u})^2$ and t_m is the appropriate value from "Students" distribution with $m - 1$ degrees of freedom. The width of the interval is $2t_m s/m^{\frac{1}{2}}$ and if this is less than the desired width d, no second step is required. If $2t_m s/m^{\frac{1}{2}} > d$, then n additional observations $w_1, w_2, \cdots, w_n$ are taken where $n \geqq (4t_m^2 s^2/d^2) - m$, and the α confidence interval is

$$\bar{z} - \frac{t_m s}{\sqrt{m+n}} \leqq \mu \leqq \bar{z} + \frac{t_m s}{\sqrt{m+n}},$$

where

$$\bar{z} = \frac{n\bar{w} + m\bar{u}}{m+n}.$$

The width of the interval is $2t_m s/(m+n)^{\frac{1}{2}}$, and this is less than d.

It is to be noted that observations in the second sample are used *only* to compute the mean, $\bar{z}$.

Let us assume that the observations in the first step are taken from a normal population with mean μ_1 and variance σ_1^2, and in the second step the mean is μ_2 and the variance σ_2^2. Stein's method is valid if $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$. However, if $\mu_1 \neq \mu_2$, but $\sigma_1^2 = \sigma_2^2$, the method can still be used to set a specified confidence interval on μ_2; the only alteration is that the second step requires a sample of size $n + m$ and $\bar{z}$ is the mean of this sample. That is to say, the sample mean from step one is not used in computing the interval. In this case if the inequality

$$\frac{4t_m^2 s^2}{d^2} \leqq n$$

in Stein's procedure is compared with the inequality

$$\frac{2t_0\sqrt{z \cdot f(n)}}{\sqrt{f_m \cdot n(n-1)}} \leqq d$$

for the method presented in this paper, it is evident that Stein's procedure is to be preferred.

Next suppose that $\mu_1 \neq \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$. Then Stein's procedure gives a confidence interval on μ_2 with *known* probability (equal to 1) of a specified *width* but the confidence coefficient is *not known*. The method presented in this paper will give a confidence interval on μ_2 with unknown probability of a specified width, but with known confidence coefficient.

Therefore, there may be cases when an experimenter would prefer the method in this paper over the one given by Stein for the mean of a normal distribution.

REFERENCE

[1] CHARLES STEIN, "A two sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

NOTES

AN EXTENSION OF THE OPTIMUM PROPERTY OF THE SEQUENTIAL PROBABILITY RATIO TEST

BY M. A. GIRSHICK[1]

Stanford University

Let $f(x, \theta)$ be a family of densities or discrete probability functions depending on the parameter θ. Let H_0 be the hypothesis $\theta = \theta_0$ and H_1 the hypothesis that $\theta = \theta_1$. A sequential probability ratio test of H_0 versus H_1 is defined by two numbers A and B. After drawing the mth observation, sampling is continued if

$$B < \prod_{i=1}^{m} \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} < A, \tag{1}$$

where $x_1, \cdots, x_m$ are the first m observations. If the probability ratio is at least equal to A, H_1 is accepted, and if it is not greater than B, H_0 is accepted.

For any sequential procedure T, let the operating characteristic be

$$L(\theta, T) = \Pr\{\text{Accepting } H_0 \mid \theta, T\}, \tag{2}$$

and let $\mathcal{E}_\theta(n \mid T)$ be the expected number of observations required by T when sampling from $f(x, \theta)$. The so-called optimum property (see [5], for instance) of a sequential probability ratio test, say T^*, is that if $L(\theta_0, T) \geqq L(\theta_0, T^*)$ and $L(\theta_1, T) \leqq L(\theta_1, T^*)$, then

$$\mathcal{E}_{\theta_0}(n \mid T) \geqq \mathcal{E}_{\theta_0}(n \mid T^*), \qquad \mathcal{E}_{\theta_1}(n \mid T) \geqq \mathcal{E}_{\theta_1}(n \mid T^*).$$

In many cases this optimum property can be extended to all values of the parameter. Suppose $\theta_0 < \theta_1$, and let $\bar{\theta}$ be a number to be defined later such that $\theta_0 < \bar{\theta} < \theta_1$. Under conditions stated below, we give the extended optimum property. If

$$\begin{aligned} L(\theta, T) &\geqq L(\theta, T^*), \qquad \theta < \bar{\theta}, \\ L(\theta, T) &\leqq L(\theta, T^*), \qquad \theta > \bar{\theta}, \end{aligned} \tag{3}$$

for all $\theta \neq \bar{\theta}$, then

$$\mathcal{E}_\theta(n \mid T) \geqq \mathcal{E}_\theta(n \mid T^*) \tag{4}$$

Received May 9, 1956; revised November 15, 1957.

[1] The result reported in this note was mentioned by the late M. A. Girshick to several of his colleagues, but was unpublished at the time of his death. Since I think the result is of sufficient interest to be in the literature, I have taken the liberty of writing this note in Girshick's name. T. W. Anderson.

The research was sponsored by the Office of Naval Research.

for all θ. Inequalities (3) indicate the premise that T is everywhere as good as T^* in the sense that the operating characteristic for T is at least as high as for T^* for θ on one side of $\bar{\theta}$ and is as least as low as for T^* on the other side of $\bar{\theta}$. Then T^* is everywhere as good as T in terms of expected number of observations.

To demonstrate the property we assume that for $\theta \neq \bar{\theta}$ there is a unique nonzero root, say $h(\theta)$, of

$$\mathcal{E}_\theta \left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^h = 1, \tag{5}$$

and that $h(\theta) > 0$ for $\theta < \bar{\theta}$ and $h(\theta) < 0$ for $\theta > \bar{\theta}$. (See [4] for discussion of the assumption and of the technique used here.) This implies that given θ_0 and θ_1 the value of $\bar{\theta}$ for which the assumption holds is unique. We make the further assumption that for each θ there is a θ' such that

$$\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} f(x, \theta) = f(x, \theta'). \tag{6}$$

We now prove (4) for $\theta < \bar{\theta}$ by assuming (3) for θ and θ'. Since

$$h(\theta') = -h(\theta),$$

we have $\theta' > \bar{\theta}$. The sequential probability ratio test T^* defined by (1) can also be defined by

$$B^{h(\theta)} < \prod_{i=1}^{m} \left[\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}\right]^{h(\theta)} < A^{h(\theta)} \tag{7}$$

or by

$$B^{h(\theta)} < \prod_{i=1}^{m} \frac{f(x_i, \theta')}{f(x_i, \theta)} < A^{h(\theta)}. \tag{8}$$

Then (4) follows by the usual optimum property because T^* is a sequential probability ratio test for testing hypothesis θ versus the hypothesis θ'. For $\theta > \bar{\theta}$ a similar argument can be used.

The conditions assumed for this extended property are satisfied by many distributions. In particular the existence of such so-called conjugate pairs for distributions of the Koopman-Darmois form has been shown [2]. Savage [3] has shown that the assumptions restrict the families to have a certain exponential form (which includes the Koopman-Darmois form). This note makes explicit Blasbalg's statement [1] that a sequential probability ratio test is optimum at an infinity of parameter points.

REFERENCES

[1] H. Blasbalg, "Sequential analysis," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1024–1028.
[2] M. A. Girshick, "Contributions to the theory of sequential analysis II, III," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 282–298.

[3] L. J. Savage, "When different pairs of hypotheses have the same family of likelihood-ratio test regions," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1028–1032.
[4] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1947.
[5] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–339.

A NOTE ON BALANCED DESIGNS

By V. R. Rao

University of Bombay

0. Summary. It is proved that a necessary and sufficient condition for a general design to be balanced is that the matrix of the adjusted normal equations for the estimates of treatment effects has $v - 1$ equal latent roots other than zero.

1. Estimates and their properties. We consider a design whose incidence matrix is $N_{v \times b} = [n_{ij}]$ in which the ith treatment is replicated r_i times and the blocks are of sizes $k_1, \cdots, k_b$. With the usual assumptions, the adjusted normal equations for the treatment effects are

$$Q = C\hat{\tau}, \tag{1.1}$$

where

$$Q = T - N \operatorname{diag}\left(\frac{1}{k_1}, \cdots, \frac{1}{k_b}\right) B \tag{1.2}$$

and

$$C = \operatorname{diag}(r_1, \cdots, r_v) - N \operatorname{diag}\left(\frac{1}{k_1}, \cdots, \frac{1}{k_b}\right) N' \tag{1.3}$$

with the condition

$$E_{1v}\hat{\tau} = 0 \tag{1.4}$$

(where E_{pq} denotes a $p \times q$ matrix with all its elements as unity).

It is well known that if rank $C = v - t$, a set of $t - 1$ independent treatment contrasts are not estimable. But if rank $C = v - 1$ every contrast is estimable and in this case the design is said to be connected.

If the design is connected there are $v - 1$ non-zero latent roots, say, $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$. As the rows of C add to zero, $(v^{-1/2}, \cdots, v^{-1/2})$ is the latent vector corresponding to the root zero.

Let

$$L = \left[\frac{L_1}{v^{-\frac{1}{2}}E_{1v}}\right] = \left[\frac{(l_{ij})}{v^{-\frac{1}{2}}E_{1v}}\right] \tag{1.5}$$

be an orthogonal matrix transforming C into diagonal form.

Received February 28, 1957; revised July 15, 1957.

Since L is orthogonal,

$$I = L'L = L_1'L_1 + \frac{1}{v} E_{vv}. \tag{1.6}$$

Pre-multiplying (1.1) by L, we get

$$LQ = LC\hat{\tau} = \left(\frac{\text{diag }(\lambda_1, \cdots, \lambda_{v-1})}{0}\right) L_1 \hat{\tau}. \tag{1.7}$$

Hence we get

$$L_1 \hat{\tau} = \text{diag}\left(\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_{v-1}}\right) L_1 Q.$$

Premultiplying by L_1' and using (1.6) and (1.4), we obtain

$$\hat{\tau} = DQ, \tag{1.8}$$

Where

$$D = [d_{ij}] = L_1' \text{ diag}\left(\frac{1}{\lambda_1}, \cdots \cdots, \frac{1}{\lambda_{v-1}}\right) L_1. \tag{1.9}$$

From the solution (1.8), it follows that

$$V(\hat{\tau}) = D\sigma^2,$$

$$\begin{aligned} V(\hat{\tau}_i - \hat{\tau}_j) &= (d_{ii} + d_{jj} - 2\, d_{ij})\sigma^2 \\ &= \sum_{\nu=1}^{v-1} \frac{(l_{\nu i} - l_{\nu j})^2}{\lambda\nu} \sigma^2. \end{aligned} \tag{1.10}$$

$$\begin{aligned} \text{Average variance} &= \frac{1}{v(v-1)} \sum_{\substack{i \\ i \neq j}} \sum_{j=1}^{v} V(\hat{\tau}_i - \hat{\tau}_j) \\ &= \frac{2\sigma^2}{v-1} \sum_{\nu=1}^{v-1} \frac{1}{\lambda\nu} \end{aligned} \tag{1.11}$$

in view of the orthogonality conditions, a result which was obtained by O. Kempthorne in an alternative way [1].

Definition. A design is said to be balanced if every elementary contrast, $\tau_i - \tau_j$ is estimated with the same variance.

2. Theorem. *A necessary and sufficient condition for a design to be balanced is that C has $v - 1$ equal latent roots other than zero.*

To prove that the condition is necessary it is enough to show that $\lambda_1 = \cdots = \lambda_{v-1}$, for the C matrix of a balanced design is of rank $v - 1$.

From (1.10) and (1.11), we get

$$\sum_{\nu=1}^{v-1} \frac{(l_{\nu i} - l_{\nu j})^2}{\lambda_\nu} = \frac{2}{v-1} \sum_{j=1}^{v-1} \frac{1}{\lambda_\nu} = \frac{1}{v-1} \sum_{k=1}^{v-1} \frac{1}{\lambda_k} \sum_{\nu=1}^{v-1} (l_{\nu i} - l_{\nu j})^2.$$

Hence

$$\sum_{\nu=1}^{v-1} (l_{\nu i} - l_{\nu j})^2 \left(\frac{1}{\lambda_\nu} - \frac{1}{v-1} \sum_{k=1}^{v-1} \frac{1}{\lambda_k}\right) = 0. \tag{2.1}$$

Consider

$$V(\hat{\tau}_j - \hat{\tau}_k) = V(\hat{\tau}_i - \hat{\tau}_j) + V(\hat{\tau}_i - \hat{\tau}_k) - 2 \operatorname{Cov}(\hat{\tau}_i - \hat{\tau}_j, \hat{\tau}_i - \hat{\tau}_k).$$

Hence,

$$\operatorname{Cov}(\hat{\tau}_i - \hat{\tau}_j, \hat{\tau}_i - \hat{\tau}_k) = \frac{1}{v-1} \sum_{\nu=1}^{v-1} \frac{1}{\lambda_\nu},$$

i.e.,

$$\sum_{\nu=1}^{v-1} \frac{(l_{\nu i} - l_{\nu j})(l_{\nu i} - l_{\nu k})}{\lambda_\nu} = \frac{1}{v-1} \sum_{\nu=1}^{v-1} \frac{1}{\lambda_\nu} \sum_{\nu'=1}^{v-1} (l_{\nu' i} - l_{\nu' j})(l_{\nu' i} - l_{\nu' k}).$$

Hence,

$$\sum_{\nu=1}^{v-1} (l_{\nu i} - l_{\nu j})(l_{\nu i} - l_{\nu k}) \left(\frac{1}{\lambda_\nu} - \frac{1}{v-1} \sum_{\nu'=1}^{v-1} \frac{1}{\lambda_{\nu'}}\right) = 0 \qquad \text{for } i \neq j \neq k, \tag{2.2}$$

From (2.1) and (2.2) taking $i = 1$, we get

$$d^{(j)\prime} \operatorname{diag}\left(\frac{1}{\lambda_1} - \frac{1}{v-1} \sum \frac{1}{\lambda_\nu}, \cdots, \frac{1}{\lambda_{v-1}} - \frac{1}{v-1} \sum \frac{1}{\lambda_\nu}\right) d^{(k)} = 0 \tag{2.3}$$

$$\text{for } j, k = 2, 3, \cdots v,$$

where $d^{(j)}$ is the column vector

$$\{l_{11} - l_{1j}, l_{21} - l_{2j}, \cdots, l_{v-11} - l_{v-1j}\}.$$

If

$$M = [d^{(2)}, d^{(3)}, \cdots, d^{(v)}],$$

then

$$M'M = \begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \cdots & & \cdots & \\ 1 & 1 & \cdots & 2 \end{bmatrix},$$

and det. $M'M = v \neq 0$. Hence $M'M$ and hence M are non-singular.

Therefore $d^{(2)}, \cdots, d^{(v)}$ are $v - 1$ linearly independent $(v - 1)$-vectors. Any $(v - 1)$ vector, say ξ, can be uniquely expressed in terms of these vectors, say

$$\xi = C_2 d^{(2)} + C_3 d^{(3)} + \cdots + C_v d^{(v)}.$$

From (2.3) it follows that

$$\xi' \operatorname{diag}\left(\frac{1}{\lambda_1} - \frac{1}{v-1}\sum\frac{1}{\lambda_r}, \cdots, \frac{1}{\lambda_{v-1}} - \frac{1}{v-1}\sum\frac{1}{\lambda_r}\right)\xi = 0$$

$$= \sum_{i,j=2}^{v} C_i C_j d^{(j')} \operatorname{diag}\left(\frac{1}{\lambda_1} - \frac{1}{v-1}\sum\frac{1}{\lambda_r}, \cdots, \frac{1}{\lambda_{v-1}} - \frac{1}{v-1}\sum\frac{1}{\lambda_r}\right)d^{(j)} = 0. \tag{2.4}$$

By taking successively ξ to be the $(v-1)$ vectors $(1, 0, \cdots, 0), \cdots$ and $(0, 0, \cdots, 1)$, we get

$$\frac{1}{\lambda_1} = \frac{1}{\lambda_2} = \cdots = \frac{1}{\lambda_{v-1}}.$$

Hence

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{v-1}.$$

The condition is sufficient, for, if rank $C = v - 1$ and $\lambda_1 = \cdots = \lambda_{v-1} = \lambda$ (say), it follows immediately that every elementary contrast is estimable, and the solutions become

$$\hat{\tau} = \frac{1}{\lambda} L_1' L_1 Q = \frac{1}{\lambda}\left(I - \frac{1}{v} E_{vv}\right) Q = \frac{Q}{\lambda}.$$

which shows that $V(\hat{\tau}_i - \hat{\tau}_j) = (2/\lambda)\sigma^2$, which is independent of both i and j and hence, the design is balanced. Q.E.D.

Corollaries. (i) *If the design is balanced, then*

$$C = \lambda I - \frac{\lambda}{v} E_{vv} \tag{2.5}$$

and the solutions are

$$\hat{\tau}_j = \frac{Q_j}{\lambda}. \tag{2.6}$$

(ii) *In a balanced design with equal block sizes, k, the replicate numbers must be equal.*

Proof. $C = \operatorname{diag}(r_1, \cdots, r_v) - 1/kNN'$ if block size is constant. Hence by Eq. (2.5), if the design is also balanced, we have

$$r_i - \frac{r_i}{k} = \lambda - \frac{\lambda}{v}.$$

Hence, r_i is the same constant for all i. Q.E.D.

(iii) *If all the treatments are replicated the same number of times and the blocks are of the same size then the only balanced design is* BIBD, *if such a design exists.*

Proof. If r is the number of replications and k is the block size, then

$$C = rI - \frac{1}{k} NN'$$

$$= \lambda I - \frac{\lambda}{v} E_{vv} \text{ by Corollary (i).} \tag{2.7}$$

Hence comparing off-diagonal elements, we get

$$\lambda_{ij} = \frac{k\lambda}{v},$$

where λ_{ij} is the number of times the pair of treatments i, j occur together in the blocks. Since λ_{ij}'s are all equal the design is Balanced Incomplete Block Design (BIBD) [2]. This result was proved in an alternative form by W. A. Thompson [3].

3. Concluding remarks. But these do not exclude the possibilities of the existence of balanced designs with different block sizes and the same number of replications. As an example consider the design whose incidence matrix is

$$N = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix},$$

$$r = (6, 6, 6, 6); \qquad k = (3, 3, 3, 3, 2, 2, 2, 2, 2, 2).$$

Here it can be verified that every elementary contrast is estimated with a variance equal to $3\sigma^2/7$, but the design is not a Balanced Incomplete Block Design.

It can also be seen that the example given above is obtained by adjoining two BIBD's with the same number of treatments. Such designs can be constructed from two BIBD's with the same number of treatments. Investigations on these lines are being carried out.

Acknowledgement. The author wishes to express his indebtedness to Professor M. C. Chakrabarti for suggesting this problem and for his help and guidance in preparing this note.

REFERENCES

[1] O. Kempthorne, "The Efficiency Factor of an Incomplete Block Design," *Ann. Math. Stat.* Vol. 27 (1956), pp. 846–849.

[2] F. Yates, "Incomplete Randomized Blocks," *Annals of Eugenics*, Vol. VII (1937), pp. 121–140.

[3] W. A. Thompson, "A Note on Balanced Incomplete Block Designs," *Ann. Math. Stat.* Vol. 27 (1956), pp. 842–846.

THE SPACING OF OBSERVATIONS IN POLYNOMIAL REGRESSION

By P. G. Guest

University of Sydney, Australia

1. Introduction and summary. De la Garza ([1], [2]) has considered the estimation of a polynomial of degree p from n observations in a given range of the

Received April 8, 1957.

independent variable x. This range may conveniently be taken to be from $+1$ to -1. He showed that for any arbitrary distribution of the points of observation there was a distribution of the n observations at only $p+1$ points for which the variances (determined by the matrix $\mathbf{X}^T \mathbf{W} \mathbf{X}$) were the same. He then considered how these $p+1$ points should be distributed so that the maximum variance of the fitted value in the range of interpolation should be as small as possible. In the present note general formulae will be obtained for the distribution of the points of observation and for the variances of the fitted values in the minimax variance case, and the variances will be compared with those for the uniform spacing case.[1]

2. Spacing for minimax variance. The fitted value is given by

$$u_p(x) = \sum_{j=0}^{p} L_j(x)\bar{y}_j, \tag{1}$$

where $L_j(x)$ is the Lagrangian coefficient corresponding to the point of observation x_j and $\bar{y}_j$ is the mean of the observed values at this point. The variance of the fitted value is $\operatorname{var} u_p(x) = \sum_{j=0}^{p} L_j^2(x) \operatorname{var} \bar{y}_j$.

At a point of observation

$$L_j(x_k) = \delta_{jk}$$

and

$$\operatorname{var} u_p(x_j) = \operatorname{var} \bar{y}_j .$$

The largest value of this variance will be as small as possible when the n observations are equally divided among the $p+1$ points. When this is done

$$\operatorname{var} u_p(x_j) = (p+1)\sigma^2/n \tag{2}$$

and

$$\operatorname{var} u_p(x) = \sum_{j=0}^{p} L_j^2(x)(p+1)\sigma^2/n. \tag{2.1}$$

Since this is a polynomial of degree $2p$, the minimax variance conditions are obtained when the maxima of $\operatorname{var} u_p(x)$ are at the $p-1$ internal points x_j, and the end points x_0 and x_p are $+1$ and -1; for then $\operatorname{var} u_p(x)$ never exceeds

$$(p+1)\sigma^2/n$$

in the range $+1$ to -1. The minimax variance conditions are thus

$$L_j'(x_j) = 0, \qquad j = 1 \text{ to } p-1. \tag{3}$$

[1] K. Smith, in an earlier discussion, has given details of curves up to the sixth degree (Biometrika 12 (1918), pp. 1-85).

Now, if

(4) $$F(x) = \prod_{j=0}^{p} (x - x_j),$$

then

$$L_j(x) = \frac{F(x)}{(x - x_j)F'(x_j)},$$

and so

$$F''(x) = \{(x - x_j)L_j''(x) + L_j'(x)\}F'(x_j),$$

and (3) is equivalent to

(5) $$F''(x_j) = 0, \qquad j = 1 \text{ to } p - 1.$$

The function $F(x)$ will be of the form $\alpha(x^2 - 1)\phi_{p-1}(x)$, where the polynomial $\phi_{p-1}(x)$ of degree $p - 1$ is determined by the $p - 1$ equations (5). The polynomial which satisfies these equations is readily shown to be the derivative $P_p'(x)$ of the Legendre polynomial. For if

(6) $$F(x) = \alpha(x^2 - 1)P_p'(x),$$

then

$$F'(x) = \alpha \frac{d}{dx}\{(x^2 - 1)P_p'(x)\} = \alpha p(p + 1)P_p(x)$$

and

$$F''(x) = \alpha p(p + 1)P_p'(x),$$

and so $F''(x)$ vanishes at the internal points $F(x) = 0$.

The points of observation for minimax variance are then to be located at $+1$, -1, and the roots of $P_p'(x) = 0$.

Since the internal points of observation are points of maximum variance, the variance will be given by an equation of the form

(7) $$\text{var } u_p(x) = \{1 + \beta(x^2 - 1)P_p'^2(x)\}(p + 1)\sigma^2/n.$$

The minima of the variance curve then occur at points for which

$$xP_p'(x) + (x^2 - 1)P_p''(x) = 0,$$

and this equation is equivalent to

(8) $$xP_p'(x) = p(p + 1)P_p(x).$$

From (2.1),

(9) $$\text{var } u_p(x) = \sum_{j=0}^{p} \left\{\frac{\alpha(x^2 - 1)P_p'(x)}{(x - x_j)\alpha p(p + 1)P_p(x_j)}\right\}^2 (p + 1)\sigma^2/n,$$

and so, on comparing the coefficients of $x^2P_p'^2(x)$ in (7) and (9),

$$\beta = \sum_{j=0}^{p} \{p(p+1)P_p(x_j)\}^{-2}.$$

The Lobatto quadrature formula [3] with $f(x) \equiv 1$ gives

$$\int_{-1}^{1} dx = \sum_{j=0}^{p} 2\{p(p+1)P_p^2(x_j)\}^{-1} = 2.$$

Thus the explicit formula for the variance of the fitted value is

$$\text{(10)} \qquad \operatorname{var} u_p(x) = \left\{ 1 + \frac{x^2 - 1}{p(p+1)} P_p'^2(x) \right\} (p+1)\sigma^2/n.$$

In the region of extrapolation, when $|x|$ is large

$$P_p'(x) \doteqdot p\{(2p)!/2^p p!^2\} x^{p-1},$$

and so

$$\text{(11)} \qquad \operatorname{var} u_p(x) \doteqdot p\{(2p)!/2^p p!^2\}^2 x^{2p} \sigma^2/n.$$

3. Uniform spacing. When the observations are spaced at equal intervals the variance of the fitted value is

$$\operatorname{var} u_p(x) = \sum_{j=0}^{p} \left\{ T_j^2(x) \Big/ \sum_{i=1}^{n} T_j^2(x_i) \right\} \sigma^2,$$

where the $T_j(x)$ are the polynomials orthogonal over the n points of observation x_i. When n is large these polynomials will approximate to multiples of the Legendre polynomials $P_j(x)$ which are orthogonal over the continuous range $+1$ to -1. Thus

$$T_j(x) \sim k_j P_j(x)$$

and

$$\sum_i T_j^2(x_i)\Delta x_i \sim k_j^2 \int_{-1}^{1} P_j^2(x)\, dx = 2k_j^2/(2j+1).$$

The interval Δx_i between neighboring observations is $2/n$, and so

$$\sum T_j^2(x_i) \sim nk_j^2/(2j+1)$$

and

$$\text{(12)} \qquad \operatorname{var} u_p(x) \sim \sum_{j=0}^{p} (2j+1) P_j^2(x) \sigma^2/n.$$

The maxima and minima of variance are at points given by

$$\sum_{j=0}^{p} (2j+1) P_j(x) P_j'(x) = 0,$$

which from the recurrence relations for Legendre polynomials is

$$\sum_{j=0}^{p} P_j'(x)\{P_{j+1}'(x) - P_{j-1}'(x)\} = 0,$$

or

$$P_p'(x)P_{p+1}'(x) = 0.$$

The points of maximum variance are then the roots of $P_p'(x) = 0$ and the points of minimum variance the roots of $P_{p+1}'(x) = 0$. It is interesting to observe that the points of observation in the minimax variance method are points of maximum variance in the uniform spacing method. These points are also the points used in the Lobatto quadrature formula.

The Christoffel-Darboux identity [3] for the sum in equation (12) leads to the alternative form

$$\text{(12.1)} \qquad \operatorname{var} u_p(x) \sim \{P_p(x)P_{p+1}'(x) - P_p'(x)P_{p+1}(x)\}(p+1)\sigma^2/n.$$

By use of the recurrence relations for the Legendre polynomials this can be put in the form

$$\text{(12.2)} \qquad \operatorname{var} u_p(x) \sim \left\{(p+1)P_p^2(x) - \frac{x^2-1}{p+1} P_p'^2(x)\right\}(p+1)\sigma^2/n.$$

At the end-points $+1$ and -1, $P_p^2(x)$ is unity and

$$\operatorname{var} u_p(\pm 1) \sim (p+1)^2\sigma^2/n.$$

At the centre of the range the variance can be obtained by substituting the values of $P_p(0)$ and $P_p'(0)$ in (12.2). It is found that

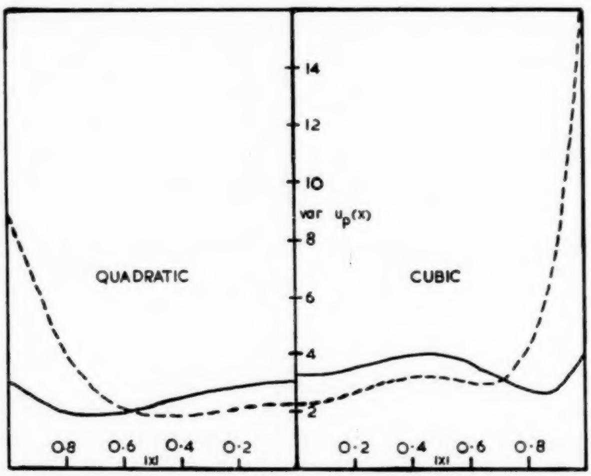


FIG. 1. The solid curve shows the variance of the fitted value for the minimax variance method and the dotted curve the variance for the uniform spacing method. The unit for the variance scale is σ^2/n.

$$\text{var } u_p(0) \sim \left\{ \frac{(2q+1)(2q-1)\cdots 1}{2^q q!} \right\}^2 \sigma^2/n, \tag{13}$$

where q is $\frac{1}{2}p$ when p is even and $\frac{1}{2}(p-1)$ when p is odd. In the region of extrapolation, when $|x|$ is large (12.2) gives

$$\text{var } u_p(x) \doteqdot (2p+1)\{(2p)!/2^p p!^2\}^2 x^{2p} \sigma^2/n.$$

The deviations from these formulae when n is not large have been discussed and tabulated [4].

4. Comparison of the two methods. In the central part of the range the uniform spacing method gives a smaller variance than the minimax variance method. An asymptotic expansion of (13) using Stirling's factorial approximation shows that the ratio of the variances is roughly $2/\pi$. This ratio increases steadily with $|x|$, and at the ends of the range the variance for the uniform spacing method exceeds that for the minimax variance method by a factor $p+1$, while in the region of extrapolation this factor approaches $2+p^{-1}$. The crossover points for the two variance curves occur at ± 0.58 for the quadratic and ± 0.72 for the cubic. Thus over most of the region of interpolation the advantage lies with the uniform spacing method, but at the extremes of the region of interpolation and in the region of extrapolation the advantage lies decidedly with the minimax variance method.

Fig. 1 shows the shape of the two variance curves in the region of interpolation for the second and third degree polynomials. Since the curves are symmetrical about the origin of x, only half of each curve is drawn.

REFERENCES

[1] A. DE LA GARZA, "Spacing of information in polynomial regression," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 123–130.

[2] A. DE LA GARZA, D. S. HAWXHURST, AND L. T. NEWMAN, "Some minimum cost experimental procedures in quadratic regression," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 178–184.

[3] F. B. HILDEBRAND, *Introduction to Numerical Analysis*, McGraw-Hill Book Co., New York, 1956, pp. 319–345.

[4] P. G. GUEST, "Estimation of the error at a point on a least-squares curve," *Australian J. Sci. Research Ser. A*, Vol. 3 (1950), pp. 173–182.

CONDITIONS THAT A STOCHASTIC PROCESS BE ERGODIC[1]

BY EMANUEL PARZEN

Stanford University

In his work on statistical inference on stochastic processes, Grenander has pointed out ([2], p. 257) that "the concept of metric transitivity seems to be

Received June 3, 1957; revised August 2, 1957.

[1] This work was supported in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

important in the problem of estimation of a stationary stochastic process." In this note, we give necessary and sufficient conditions in terms of characteristic functions that a strictly stationary stochastic process $X(t)$ be metrically transitive or ergodic (see Doob [1], pp. 452–457 for definition of the terminology). More importantly, we state a mean ergodic theorem (or weak law of large numbers) for stochastic processes which are strictly stationary of order K, by which is meant that for every choice of K points $t_1, \cdots, t_K$, the random variables $X(t_1 + h), \cdots, X(t_K + h)$ have a joint probability distribution which does not depend on h.

THEOREM 1. Let the random variables $X(t)$ be defined for t in

$$T = \{0, \pm 1, +2, \cdots\}.$$

Let K be a positive integer. Let $t_1, \cdots, t_K$ be points in T. Assume that there is a characteristic function $\varphi(u_1, \cdots, u_K)$ such that, for all $u_1, \cdots, u_K$,

$$E[\exp i\{u_1X(t_1 + h) + \cdots + u_KX(t_K + h)\}] = \varphi(u_1, \cdots, u_K) \quad \text{for all } h \text{ in } T. \tag{1.1}$$

Assume that, for each τ in T, there is a characteristic function $\varphi(u_1, \cdots, u_K; \tau)$ such that

$$\begin{aligned} E[\exp i\{u_1(X(t_1 + h) - X(t_1 + h + \tau)) + \cdots + u_K(X(t_K + h) \\ - X(t_K + h + \tau))\}] = \varphi(u_1, \cdots, u_K; \tau) \quad \text{for all } h \text{ in } T. \end{aligned} \tag{1.2}$$

Let $r \geqq 1$. Then for every Borel function $g(x_1, \cdots, x_K)$ such that

$$E\,|\,g(X(t_1), \cdots, X(t_K))\,|^r < \infty,$$

the sample means

$$M_n(g) = \frac{1}{n+1} \sum_{h=0}^{n} g(X(t_1 + h), \cdots, X(t_K + h)) \tag{1.3}$$

converge as a limit in r-mean. A necessary and sufficient condition that the limit of the $M_n(g)$ be the ensemble mean $E(g) = Eg(X(t_1), \cdots, X(t_K))$ is that, for all real $u_1, \cdots, u_K$,

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\tau=0}^{n} \varphi(u_1, \cdots, u_K; \tau) = |\,\varphi(u_1, \cdots, u_K)\,|^2. \tag{1.4}$$

The meaning of these conditions is as follows: (1.1) states, in terms of characteristic functions, that the stochastic process is strictly stationary of order K; (1.2) states that the process of increments $Y(t) = X(t) - X(t + \tau)$ is strictly stationary of order K; (1.4) represents a very weak form of asymptotic independence.

From Theorem 1, together with the Birkhoff-Khintchine ergodic theorem (see Doob [1], pp. 464–473) we immediately obtain the following theorem.

THEOREM 2: A strictly stationary stochastic process $X(t)$ is metrically transitive if, and only if, for every positive integer K, for any choice of K points $t_1, \cdots, t_K$, and for any real numbers $u_1, \cdots, u_K$, (1.4) holds.

The conditions of Theorem 2 constitute a formulation in terms of characteristic functions of known conditions for metric transitivity (see Loève [4], p. 435).

As an indication of the power of these theorems, let us mention that with their aid one can readily establish the following statement made without proof in the book of Grenander and Rosenblatt ([3], p. 44): If $X(t)$ is a normal process, a necessary and sufficient condition for it to be ergodic (metrically transitive) is that its spectrum be continuous. If $X(t)$ is a linear process, then it is ergodic.

Theorem 1, and consequently Theorem 2, may be extended to the case of continuous parameter stochastic processes. They provide a new proof of the theorem of Maruyama (see [2], p. 257) that a continuous stationary normal process is metrically transitive if, and only if, its spectrum is continuous.

Theorem 1 is very closely related to the weak law of large numbers for wide-sense stationary processes (see Doob [1], p. 489), from which it differs in that it does not require existence of second moments for $X(t)$.

The proof of Theorem 1 is fairly immediate. From (1.1), (1.2), and (1.4), it follows (either by the weak law of large numbers for wide-sense stationary processes, or directly by a simple argument [6]) that the theorem holds for trigonometric polynomials $g(x_1, \cdots, x_K) = \exp i(u_1x_1 + \cdots + u_Kx_K)$. To extend the theorem to Borel functions $g(x_1, \cdots, x_K)$ such that $E \mid g \mid^r < \infty$, one uses the fact that to any $\epsilon > 0$ one may find a trigonometric polynomial $g_\epsilon(x_1, \cdots, x_K)$ such that

$$E \mid g(X(t_1), \cdots, X(t_K)) - g_\epsilon(X(t_1), \cdots, X(t_K)) \mid^r < \epsilon.$$

In [5] one may find related theorems, including a discussion of convergence with probability one of certain sample means $M_n(g)$ of stochastic processes which are strictly stationary of order K.

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, New York, 1953.

[2] U. Grenander, "Stochastic processes and statistical inference," *Arkiv for Matematik*, Vol. 1 (1950), pp. 195–277.

[3] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, New York, 1957.

[4] M. Loève, *Probability Theory*, New York, 1955.

[5] E. Parzen, "Conditions that a stochastic process be ergodic," Stanford University, Department of Statistics, Technical Report No. 11, March, 1957.

[6] E. Parzen, "Laws of large numbers for dependent random variables," to be published as a section in a forthcoming text entitled *An Introduction to Modern Probability Theory*.

POWER FUNCTIONS OF THE GAMMA DISTRIBUTION

By Gerald D. Berndt

Advisory Committee on Weather Control

Power functions are given for testing hypotheses on an increase in the mean μ of a gamma variable.

Let x be a random variable from a gamma population and let the frequency distribution of x be given by

$$\begin{aligned} f_0 = f(x; \beta, \gamma) &= (\beta^\gamma \Gamma(\gamma))^{-1} x^{\gamma-1} \exp(-x/\beta), & x > 0, \\ &= 0, & x \leqq 0, \end{aligned}$$

where $\beta > 0$ and $\gamma > 0$. If x then undergoes a scale change of the form $x \to \delta x$ with $\delta > 1$, it is easily verified that the frequency distribution of δx is given by

$$\begin{aligned} f_1 = f(\delta x; \delta\beta, \gamma) &= ((\delta\beta)^\gamma \Gamma(\gamma))^{-1} x^{\gamma-1} \exp(-x/\delta\beta), & x > 0, \\ &= 0, & x \leqq 0. \end{aligned}$$

Now in testing the null hypothesis $H_0 : \mu = \beta\gamma$ against the alternative hypothesis $H_1 : \mu = \delta\beta\gamma$, $\delta > 1$ and specified, the probability of detecting the hypothesized change in the mean, or the power of the test, is given by

$$\pi_\delta = \int_{x(\alpha)}^{\infty} f_1 \, dx,$$

where $x(\alpha)$ is such that

$$\alpha = \int_{x(\alpha)}^{\infty} f_0 \, dx$$

and α is the significance level of the test.

Curves of power functions of testing H_0 against H_1 are given for

$$\gamma = \tfrac{1}{2}, 1(1)5, 7, 10(5)50,$$

$$1.0 \leqq \delta \leqq 4.0,$$

and

$$\alpha = 0.01, 0.05, \text{ and } 0.10.$$

For sufficiently large γ, the distribution of x converges to the normal distribution, and for many purposes the power of the test may then be evaluated by simply using the tables of the normal distribution function with standardized variates

$$t_\alpha = (x(\alpha) - \beta\gamma)/\beta\sqrt{\gamma},$$

Received June 10, 1957.

which is exceeded with probability α under H_0, and

$$t_\pi = (t_\alpha \beta\sqrt{\gamma} + \beta\gamma(1 - \delta))/\delta\beta\sqrt{\gamma},$$

which is exceeded with probability π under H_1. The upper bound of the error in using the normal approximation to the gamma distribution for $\gamma \geqq 50$ is calculated, by trial, to be

$$\sup_x | G(x) - N(x) | < 0.019,$$

where $G(x)$ is the distribution function of x as a gamma variable and $N(x)$ is the distribution function of x as a normal variable.

Consider the following example of the use of the accompanying power curves. In illustrating the use of the power curves we first take note of a well known property of the gamma distribution. That is, if $x_i(i = 1, \cdots, n)$ are independent random variables from gamma populations with parameters β and γ_i, then the sample mean is also a gamma variable with parameters β/n and

$$\gamma = \textstyle\sum_{i=1}^{n} \gamma_i .$$

See, for example, [1].

Suppose, for illustration, that a sample of size $n = 10$ is drawn, and that the $x_i(i = 1, \cdots, 10)$ are known to be independently and identically distributed gamma variables with $\gamma_i = 2.0$ and the same β for each i. It is desired to test $H_0{:}\mu = \mu_0$ against $H_1{:}\mu = 1.5\mu_0 = \mu_1$ with probability $\alpha = 0.05$ of accepting H_1 when in fact H_0 is true. What is the probability of detecting μ_1? Here $\delta = 1.5$ and $\gamma = 20.0$. In Fig. 2 we find $\delta = 1.5$ on the abscissa and move vertically to the point of intersection with the curve $\gamma = 20.0$. The power, $\pi_{1.5} = 0.598$, is then the ordinate value at this point of intersection.

How large a sample should be drawn in order that there is at least a probability of $\pi_{1.5} = 0.75$ of detecting the specified increase $\delta = 1.5$ in μ_0? Interpolating for the value of γ at $\delta = 1.5$ and $\pi_{1.5} = 0.75$, we find $\gamma = 32$. Hence the sample size should be at least $n = 32/\gamma_i = 16$ in this case.

The calculations on which the power curves are based were made using 3-point Lagrangian interpolation in Pearson's tables of the incomplete gamma function [2]. All calculations have been verified by actual integration of the gamma functions using high-speed computing machinery. This verification was carried out under the supervision of Dr. Max A. Woodbury at New York University.

I wish to acknowledge the work of Dr. Woodbury and his staff in making these calculations, and also to thank Elaine Berndt who performed all of the interpolations we have required and who drafted the accompanying figures.

REFERENCES

[1] C. R. RAO, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.

[2] *Tables of the Incomplete Gamma Function*, Karl Pearson, editor, Cambridge University Press, re-issued in 1951.

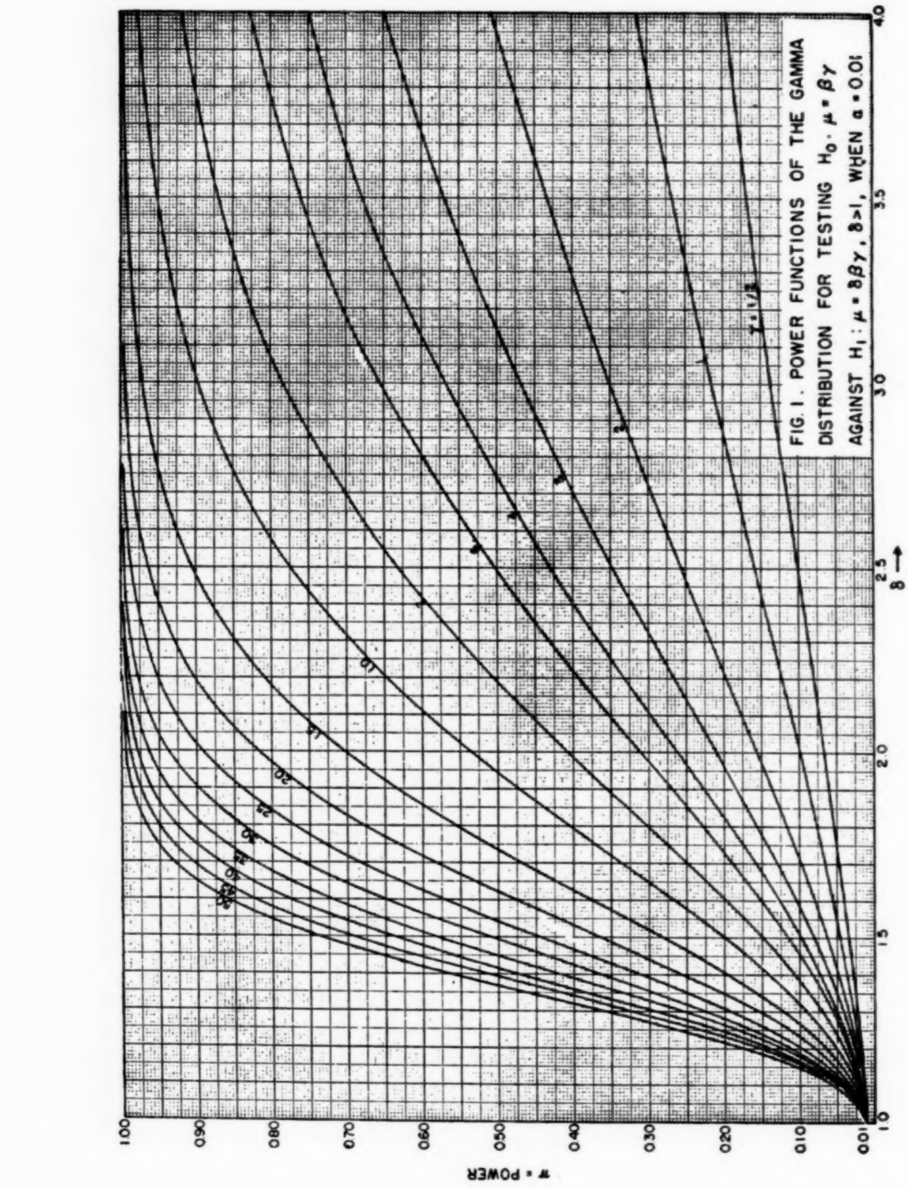


FIG. 1. POWER FUNCTIONS OF THE GAMMA DISTRIBUTION FOR TESTING $H_0: \mu = \beta\gamma$ AGAINST $H_1: \mu = \delta\beta\gamma$, $\delta > 1$, WHEN $\alpha = 0.01$

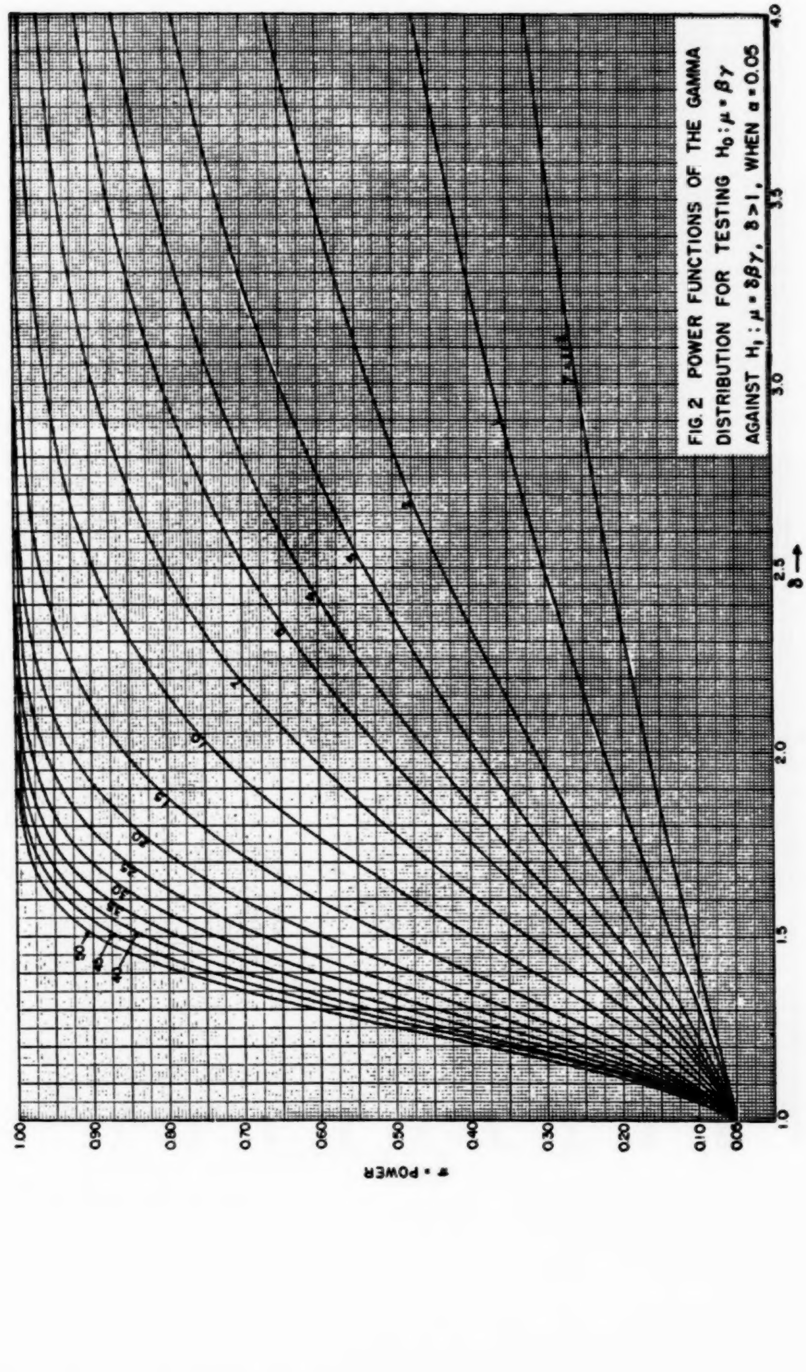


FIG. 2 POWER FUNCTIONS OF THE GAMMA DISTRIBUTION FOR TESTING $H_0: \mu = \beta\gamma$ AGAINST $H_1: \mu = \delta\beta\gamma$, $\delta > 1$, WHEN $\alpha = 0.05$

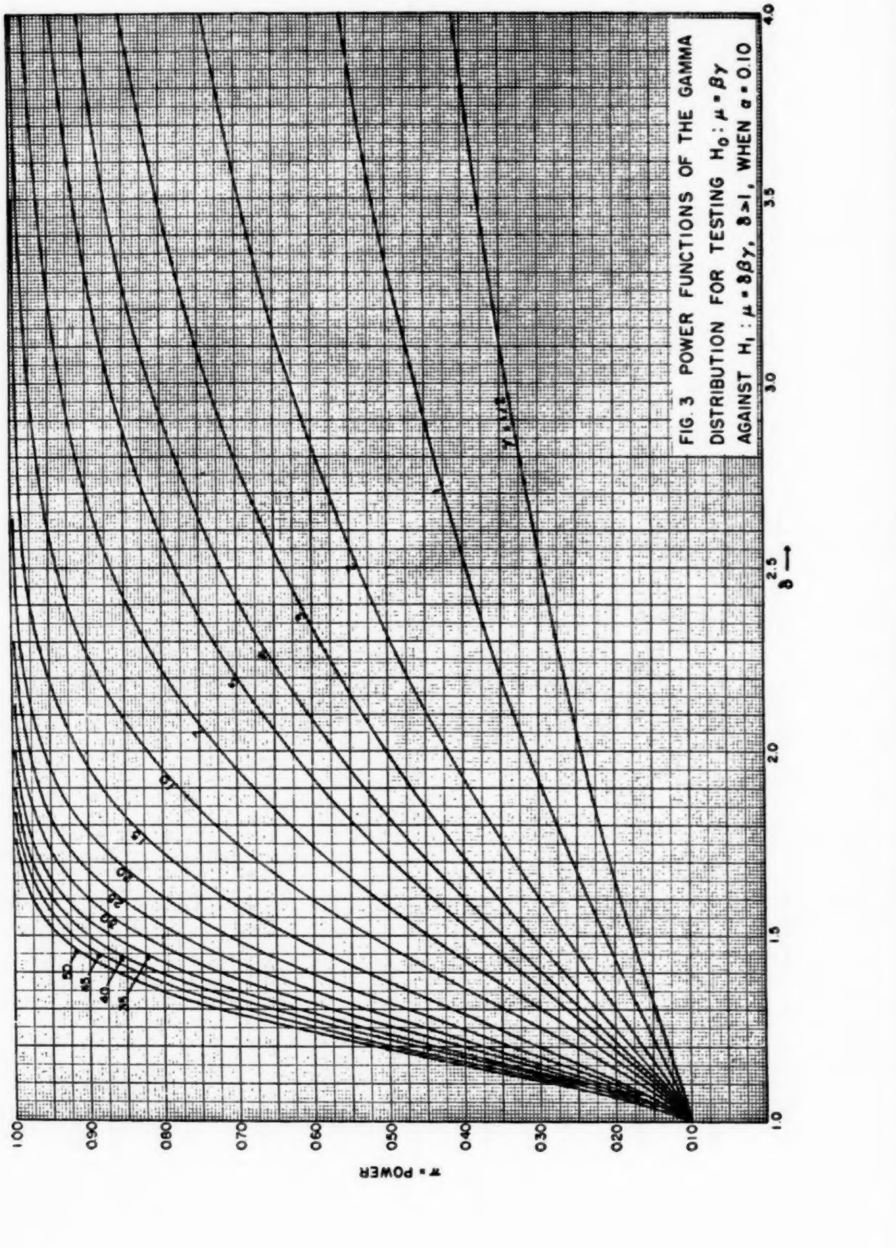


FIG. 3 POWER FUNCTIONS OF THE GAMMA DISTRIBUTION FOR TESTING $H_0: \mu = \beta\gamma$ AGAINST $H_1: \mu = \delta\beta\gamma$, $\delta > 1$, WHEN $\alpha = 0.10$

THE SMALL SAMPLE DISTRIBUTION OF $n\omega_n^2$

By A. W. Marshall

The RAND Corporation

The asymptotic distribution of the statistic

$$(1) \qquad n\omega_n^2 = n \int_{-\infty}^{\infty} [s_n(x) - F(x)]^2 \, dF(x),$$

where $S_n(x)$ is the sample cumulative distribution function (CDF), and $F(x)$ the true CDF, is known and tabled [1]. Below are tabled some values of the CDF's of $n\omega_n^2$ for $n = 1, 2$, and 3. Convergence to the asymptotic distribution appears to be extremely rapid.

1. General considerations. It is well-known that: (A) the distribution of $n\omega_n^2$ is distribution free so that it is sufficient to treat the case where $F(x)$ is uniform on the interval [0, 1]; and (B) an equivalent form, especially suitable for computation from the ordered sample $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, is

$$(2) \qquad n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - F(x_i)\right]^2$$

or for the case where $F(x)$ is uniform [0, 1]

$$(3) \qquad n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - x_i\right]^2.$$

As was suggested to me several years ago by Oliver Gross (3) clearly shows that the CDF of the $n\omega_n^2$ statistic can be evaluated rather easily for small n. The case $n = 1$ is trivial. For $n = 2$ one must evaluate the area in the intersection of a circle with its center at $x_1 = \frac{1}{4}$, $x_2 = \frac{3}{4}$ and a triangle with vertices at (0, 0), (0, 1), and (1, 1). For $n = 3$ one must evaluate the volume in the intersection of a sphere with center at the point $(\frac{1}{6}, \frac{1}{2}, \frac{5}{6})$ and the tetrahedron with vertices at (0, 0, 0), (0, 0, 1), (0, 1, 1), (1, 1, 1). From (3) one also derives the result that $n\omega_n^2$ has a minimum value of $1/12n$ and a maximum value of $n/3$.

2. Case A: $n = 1$. Since

$$\omega_1^2 = \tfrac{1}{12} + [\tfrac{1}{2} - x_1]^2$$

the CDF of ω_1^2 is

$$F_1(z) = \Pr[\omega_1^2 \leqq z] = \begin{cases} 0, & z < \frac{1}{12}, \\ (4z - \frac{1}{3})^{\frac{1}{2}}, & \frac{1}{12} \leqq z \leqq \frac{1}{3}, \\ 1, & z > \frac{1}{3}. \end{cases}$$

3. Case B: $n = 2$.

$$2\omega_2^2 = \tfrac{1}{24} + [\tfrac{1}{4} - x_1]^2 + [\tfrac{3}{4} - x_2]^2.$$

Received June 25, 1957.

By evaluating the area common to a circle of radius $(z - 1/24)^{\frac{1}{2}}$ with center at $(\frac{1}{4}, \frac{3}{4})$ and the triangle with vertices at (0, 0), (0, 1), and (1, 1), multiplying by two, the CDF $F_2(z)$ of the associated value of $2\,\omega_2^2$ is obtained. The result is:

$$
\begin{array}{ll}
0, & z < \dfrac{1}{24}, \\[2ex]
2\pi\left(z - \dfrac{1}{24}\right), & \dfrac{1}{24} \leqq z \leqq \dfrac{5}{48}, \\[2ex]
\left(z - \dfrac{1}{24}\right)\left[2\pi - 4\,\mathrm{Cos}^{-1}\,\tfrac{1}{4}(z - 1/24)^{-\frac{1}{2}} + \dfrac{(z - 5/48)^{\frac{1}{2}}}{z - 1/24}\right], & \dfrac{5}{48} < z \leqq \dfrac{1}{6}, \\[2ex]
\left(z - \dfrac{1}{24}\right)\left[\dfrac{3\pi}{2} - 2\,\mathrm{Cos}^{-1}\,\dfrac{\tfrac{1}{4}(1/8)^{\frac{1}{2}} - (z - 1/6)^{\frac{1}{2}}(z - 5/40)^{\frac{1}{2}}}{z - 1/24}\right] & \\[2ex]
\qquad + \dfrac{1}{8} + \left[\dfrac{1}{2}(z - 1/6)\right]^{\frac{1}{2}} + \tfrac{1}{2}(z - 5/48)^{\frac{1}{2}}, & \dfrac{1}{6} < z \leqq \dfrac{2}{3}, \\[2ex]
1, & z > \dfrac{2}{3}.
\end{array}
$$

4. Case C: $n = 3$. This is the first complicated case and reduces to the problem of evaluating the volume of the intersection of a sphere of radius

$$(z - 1/36)^{\frac{1}{2}},$$

with its center at $(\frac{1}{6}, \frac{1}{2}, \frac{5}{6})$, and a tetrahedron with vertices at (0, 0, 0), (0, 0, 1), (0, 1, 1), and (1, 1, 1). Whereas in the case $n = 2$ there are five intervals over which $F_2(z)$ is separately defined, when $n = 3$ there are eight: $(-\infty, 1/36)$,

TABLE I

Values of the CDF's of $n\omega_n^2$ for $n = 1, 2, 3$ and the asymptotic distribution at selected points

z	$F_1(z)$	$F_2(z)$	$F_3(z)$	$F(z)$
.11888	.37708	.46692	.47343	.50000
.14663	.50318	.57614	.57683	.60000
.16385	.56751	.63384	.63009	.65000
.18433	.63560	.68842	.68521	.70000
.20939	.71009	.73974	.74191	.75000
.24124	.79475	.79126	.79924	.80000
.28406	.89605	.84515	.85481	.85000
.34730	1.00000	.90296	.90617	.90000
.40520	1.00000	.94007	.93661	.93000
.46136	1.00000	.96554	.95723	.95000
.54885	1.00000	.98968	.97793	.97000
.74346	1.00000	1.00000	.99680	.99000
1.16786	1.00000	1.00000	1.00000	.99900

(1/36, 1/18), (1/18, 1/12), (1/12, 1/9), (1/9, 5/24), (5/24, 11/36), (11/36, 1) and $(1, \infty)$. Partial results for these intervals are as follows, where $\zeta = z - 1/36$: $F_3^{(1)}(z) = 0$; $F_3^{(2)}(z) = 8\pi\zeta^{3/2}$; $F_3^{(3)}(z) = \frac{2}{3}\pi(3z - 1/9)$; $F_3^{(4)}(z) = \frac{2}{3}\pi(3z - 1/9) - 2\pi[4\zeta^{3/2} - 2^{1/2}\zeta + 2^{-1/2}/27] + 6\zeta^{3/2}\, V(1, \zeta^{-1/2}/6)$; $\cdots$; $F_3^{(8)}(z) = 1$, where $V(1, a)$ is the volume of the wedge-shaped segment of the sphere of unit radius, center at the origin, cut out by the two planes $x = a$ and $y = a$. It is possible to obtain expressions in closed form for $F_3(z)$ over all of the eight intervals; however their derivation is tedious and the expressions complicated.[1] A numerical evaluation was therefore undertaken by the RAND Numerical Analysis section. The result of these computations are shown in Table 1 along with the calculated values of $F_n(z)$ for $n = 1$ and 2, and for the asymptotic distribution. The values of $F_3(z)$ appear to be off by one in the fifth decimal place. The rapid convergence to the asymptotic distribution, especially in the more interesting region of the tail of distribution, seems clear.

One other piece of evidence, although of a much weaker sort, is available that suggests that the asymptotic distribution is a good approximation to the exact distribution for small n. A sample of 400 values of $n\omega_n^2$ was produced for the case $n = 10$. Grouping into twenty cells using the 5th, 10th, 15th, $\cdots$, percentage points of the asymptotic distribution gave the following cell entries: 13, 19, 20, 18, 11, 21, 16, 18, 28, 17, 21, 22, 26, 18, 21, 16, 23, 25, 23, 24. Application of the χ^2 test gives a value of $\chi^2 = 17.5$. With 19 d.f. this value is exceeded with probability of approximately .55. Application of the Kolmogorov test statistic, Sup $| S_n(x) - F(x) |$, to the grouped data (for an approximate test) gives a value of 1.20. This value would be exceeded on the order of 11 per cent of the time under the null hypothesis.

Anderson and Darling in one of their papers [2] mention that "empirical study suggests that the asymptotic value is reached very rapidly, and it appears safe to use the asymptotic value for a sample size as large as 40." The results given above suggest the sample size for which it is reasonable to use the asymptotic distribution is likely to be more nearly 3 or 4, or perhaps 5.

For an allied form of the ω^2 test criterion, denoted by W_n^2 in [2] and formed by adding to (1) the weight function $\psi(X) = [F(X)(1 - F(X))]^{-1}$, an even more rapid convergence seems to occur. $F_1(z) = (1 - 4e^{-z-1})^{1/2}$ for the statistic W_1^2. Evaluating $F_1(z)$ at the 90, 95, and 99 asymptotic percentage points given in [2] yields .88716, .93292, and .98433 respectively.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 765–769.

---

[1] D. Anderson may publish the complete results later.

LIMITING DISTRIBUTIONS OF HOMOGENEOUS FUNCTIONS OF SAMPLE SPACINGS[1]

By Lionel Weiss

Cornell University

1. Summary. Suppose $T_1, T_2, \cdots, T_n$ are the lengths of n subintervals into which the interval $[0, 1]$ is broken by $(n - 1)$ independent chance variables, each with a uniform distribution on $[0, 1]$. Moran [1], Kimball [2], and Darling [3] have shown that if r is a positive number, then the asymptotic distribution of $T_1^r + T_2^r + \cdots + T_n^r$ is normal. It is the purpose of this note to extend this result in two directions: more general functions of $T_1, \cdots, T_n$ are handled, and the joint distribution of several such functions is discussed. The proof is short and very simple.

2. Notation and assumptions. As already indicated, $T_1, T_2, \cdots, T_n$ are the n subintervals into which the unit interval is randomly broken. $U_1, U_2, \cdots, U_n$ are independent chance variables, each with the density function e^{-u} for $u \geqq 0$, zero for $u < 0$. $S_n = U_1 + U_2 + \cdots + U_n$. $V_i = U_i/S_n$ for $i = 1, \cdots, n$. It is known (and is very easily verified) that S_n is distributed independently of $(V_1, V_2, \cdots, V_n)$, and that the joint distribution of

$$(V_1, V_2, \cdots, V_n)$$

is exactly the same as the joint distribution of $T_1, T_2, \cdots, T_n$.

We are given k sequences of functions:

$$\{G_{1,n}(U_1, U_2, \cdots, U_n)\}, \cdots, \{G_{k,n}(U_1, U_2, \cdots, U_n)\},$$

$n = 1, 2, \cdots$. These functions are assumed to satisfy the following conditions:

(1) $G_{i,n}(U_1, \cdots, U_n)$ is homogeneous of order r_i for all n, r_i a positive quantity;

(2) the joint distribution of

$$\frac{G_{1,n}(U_1, \cdots, U_n) - A_1 n}{B_1\sqrt{n}}, \cdots, \frac{G_{k,n}(U_1, \cdots, U_n) - A_k n}{B_k\sqrt{n}}$$

approaches a k-variate normal distribution with zero means and covariance matrix C, say, as n increases. $A_1, \cdots, A_k$ and $B_1, \cdots, B_k$ are positive constants. (The results hold for any values of $A_1, \cdots, A_k$. The assumption that they are positive is merely a convenience.)

We denote the element of C in row i and column j by c_{ij}.

3. The asymptotic distribution of $G_{1,n}(T_1, \cdots, T_n), \cdots, G_{k,n}(T_1, \cdots, T_n)$.

Theorem. *Under the assumptions of Sec. 2, the joint distribution of*

$$\frac{n^{r_1} G_{1,n}(T_1, \cdots, T_n) - A_1 n}{B_1\sqrt{n}}, \cdots, \frac{n^{r_k} G_{k,n}(T_1, \cdots, T_n) - A_k n}{B_k\sqrt{n}}$$

Received July 8, 1957.

[1] Research supported by the Office of Naval Research.

approaches a k-variate normal distribution with zero means and covariance matrix

$$\left\{c_{ij} - \frac{r_i r_j A_i A_j}{B_i B_j}\right\}$$

as n increases.

PROOF. By assumption, the distribution of the k-dimensional vector $\bar{V}(n)$ whose ith element is

$$\frac{G_{i,n}(U_1, \cdots, U_n) - A_i n}{B_i\sqrt{n}}$$

approaches the k-variate normal distribution with zero means and covariance matrix C. We rewrite the ith term of $\bar{V}(n)$ as

$$\frac{G_{i,n}(U_1, \cdots, U_n) - S_n^{r_i} A_i n^{1-r_i} + S_n^{r_i} A_i n^{1-r_i} - A_i n.}{B_i\sqrt{n}}$$

Now S_n/n converges stochastically to one as n increases; therefore the distribution of the k-dimensional vector $\bar{V}'(n)$ whose ith element is

$$\frac{G_{i,n}(U_1, \cdots, U_n) - S_n^{r_i} A_i n^{1-r_i} + S_n^{r_i} A_i n^{1-r_i} - A_i n}{\left(\frac{S_n}{n}\right)^{r_i} B_i\sqrt{n}}$$

approaches the k-variate normal distribution with zero means and covariance matrix C. $\bar{V}'(n)$ may be written as the sum of two vectors, $\bar{V}_1(n)$ and $\bar{V}_2(n)$, whose ith elements are respectively

$$\frac{n^{r_i} G_{i,n}(V_1, \cdots, V_n) - A_i n}{B_i\sqrt{n}}$$

and

$$\frac{A_i n - n^{r_i+1} A_i S_n^{-r_i}}{B_i\sqrt{n}}$$

We note that $\bar{V}_1(n)$ and $\bar{V}_2(n)$ are distributed independently of each other.

Next we examine the distribution function, say $F_n(x_1, \cdots, x_k)$, of $\bar{V}_2(n)$.

$$F_n(x_1, \cdots, x_k) = \Pr\left[\frac{A_i n - n^{r_i-1} A_i S_n^{-r_i}}{B_i\sqrt{n}} \leqq x_i ; i = 1, \cdots, k\right]$$

$$= \Pr\left[\frac{S_n - n}{\sqrt{n}} \leqq \sqrt{n}\left\{\left(\frac{A_i n}{A_i n - \sqrt{n}\, B_i x_i}\right)^{\frac{1}{r_i}} - 1\right\},\quad i = 1, \cdots k\right].$$

As n increases, the distribution of $(S_n - n)/\sqrt{n}$ approaches the standard normal distribution, by the univariate central-limit theorem. And for any fixed x_i,

$$\sqrt{n}\left\{\left(\frac{A_i n}{A_i n - \sqrt{n}\, B_i x_i}\right)^{\frac{1}{r_i}} - 1\right\} \to \frac{B_i x_i}{r_i A_i}$$

as n increases. Thus, if Z denotes a chance variable with a standard normal distribution, $F_n(x_1, \cdots, x_k)$ approaches

$$\Pr\left[\frac{r_i A_i Z}{B_i} \leqq x_i\,; i = 1, \cdots, k\right]$$

for each vector $(x_1, \cdots, x_k)$.

Next, we denote by $\rho_{1,n}(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}_1(n)$, by $\rho_{2,n}(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}_2(n)$, and by $\rho_n(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}'(n)$.

We have $\rho_n(t_1, \cdots, t_k) = \rho_{1,n}(t_1, \cdots, t_k)\cdot\rho_{2,n}(t_1, \cdots, t_k)$, or

$$\rho_{1,n}(t_1, \cdots, t_k) = \frac{\rho_n(t_1, \cdots, t_k)}{\rho_{2,n}(t_1, \cdots t_k)}.$$

As n increases,

$$\rho_n(t_1, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{k}\sum c_{ij} t_i t_j\right\}$$

and

$$\rho_{2,n}(t_1, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{k} \frac{t_j r_j A_j}{B_j}\right]^2\right\}.$$

Therefore, as n increases,

$$\rho_{1,n}(t_i, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{k}\sum t_i t_j\left[c_{ij} - \frac{r_i r_j A_i A_j}{B_i B_j}\right]\right\}.$$

This proves the theorem.

REFERENCES

[1] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), pp. 92–98.

[2] B. F. Kimball, "Some basic theorems for developing tests of fit for the case of non-parametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[3] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

ANOTHER COUNTABLE MARKOV PROCESS WITH ONLY INSTANTANEOUS STATES

By David Blackwell[1]

University of California, Berkeley

Let P be the transition function for a Markov process with a countable state space A and stationary transition probabilities; i.e., P is a nonnegative function defined for all triples (a, b, t) with $a \varepsilon A$, $b \varepsilon A$, and t a nonnegative real number, satisfying

$$P(a, b, 0) = 1 \text{ if } a = b, \qquad 0 \text{ if } a \neq b, \tag{1}$$

$$\sum_b P(a, b, t) = 1 \quad \text{for all} \quad a, t, \tag{2}$$

and

$$P(a, b, s + t) = \sum_{c \varepsilon A} P(a, c, s)P(c, b, t) \quad \text{for all} \quad s \geqq 0, \qquad t \geqq 0, a, b. \tag{3}$$

We shall suppose, as usual, that P is continuous at $t = 0$; i.e.,

$$P(a, a, t) \to 1 \text{ as } t \to 0 \text{ for all } a. \tag{4}$$

It is well known that, for any P satisfying (1), (2), (3), and (4), $P'(a, a, 0)$ exists for all a (it may be negatively infinite). Following P. Lévy [2], a state is called "instantaneous" if $P'(a, a, 0) = -\infty$. Examples of processes with all states instantaneous have been given by Feller and McKean [2] and by Dobrushin [1]. The purpose of this note is to describe a third example, somewhat simpler than those previously given.

We first describe the process informally, after which we define P and verify (1), (2), (3), and (4) and $P'(a, a, 0) = -\infty$ for all a directly. Let $X_1(t)$, $X_2(t)$, $\cdots$ be a sequence of Markov processes, independent of each other, each with two states 0 and 1. We suppose $X_n(0) = 0$ for all n. Let $X_n(t)$ be characterized by the parameters λ_n, μ_n :

$$\Pr\{X_n(t + h) = 1 \mid X_n(t) = 0\} = \lambda_n h + o(h),$$

$$\Pr\{X_n(t + h) = 0 \mid X_n(t) = 1\} = \mu_n h + o(h).$$

Our process $X(t)$ will be the joint process $X_1(t)$, $X_2(t)$, $\cdots$ which is clearly a Markov process. To insure that $X(t)$ has only a countable set of states, we

Received July 11, 1957.

[1] This paper was supported in part by funds provided under Contract AF-41(657)-29 with the Air Research and Development Command, USAF School of Aviation Medicine, Randolph Field, Texas.

determine λ_n, μ_n so that, at each time t, with probability 1, $X_n(t) = 0$ for all but a finite number of n. Since

$$\Pr(X_n(t) = 0 \mid X_n(0) = 0) = \frac{\mu_n}{\mu_n + \lambda_n} + \frac{\lambda_n}{\mu_n + \lambda_n} e^{-(\lambda_n + \mu_n)t}$$

$$\geq \frac{\mu_n}{\mu_n + \lambda_n},$$

this will occur if

$$\prod_n \frac{\mu_n}{\mu_n + \lambda_n} > 0, \tag{5}$$

i.e.,

$$\sum_n \frac{\lambda_n}{\lambda_n + \mu_n} < \infty.$$

A state is instantaneous if and only if the probability of remaining in it throughout an interval is zero. Since the probability that $X_n(t) = 0$ throughout T, $T + h$ given that $X_n(T) = 0$ is $e^{-\lambda_n h}$, the chance that the state $X(T)$ with $X_n(T) = 0$ for $n \geq N$ will persist throughout T, $T + h$ is at most

$$\prod_N^\infty e^{-\lambda_n h} = e^{-h(\lambda_N + \lambda_{N+1} + \cdots)},$$

and will be zero if

$$\sum_n \lambda_n = \infty. \tag{6}$$

Thus any choice of $\{\lambda_n\}$, $\{\mu_n\}$ satisfying (5) and (6) yields an example of a process with only instantaneous states.

Formally, the set A of states is the set of all infinite sequences

$$a = (\epsilon_1, \epsilon_2, \cdots)$$

of 0's and 1's with only finitely many 1's. Let $\{\lambda_n\}$, $\{\mu_n\}$ be sequences of positive numbers satisfying (5) and (6), let

$$R_n(0, 0, t) = \frac{\mu_n}{\mu_n + \lambda_n} + \frac{\lambda_n}{\mu_n + \lambda_n} e^{-(\lambda_n + \mu_n)t},$$

$$R_n(1, 1, t) = \frac{\lambda_n}{\mu_n + \lambda_n} + \frac{\mu_n}{\mu_n + \lambda_n} e^{-(\lambda_n + \mu_n)t},$$

$$R_n(0, 1, t) = 1 - R_n(0, 0, t),$$

$$R_n(1, 0, t) = 1 - R_n(1, 1, t),$$

and define, for any two states $a = (\epsilon_1, \epsilon_2, \cdots)$ and $b = (\delta_1, \delta_2, \cdots)$ and any $t \geqq 0$,

$$P(a, b, t) = \prod_{n=1}^{\infty} R_n(\epsilon_n, \delta_n, t). \tag{7}$$

Denote by A_N the set of all states $a = (\epsilon_1, \epsilon_2, \cdots)$ with $\epsilon_n = 0$ for all $n > N$. For $a \in A_N$ and any $M \geqq N$, we have

$$\sum_{b \in A_M} P(a, b, t) = h_M(t) \sum_{\delta_1, \cdots, \delta_M} \prod_1^M R_n(\epsilon_n, \delta_n, t)$$

$$= h_M(t) \prod_1^M (R_n(\epsilon_n, 0, t) + R_n(\epsilon_n, 1, t)) = h_M(t),$$

where

$$h_M(t) = \prod_{M+1}^{\infty} R_n(0, 0, t) \geqq \prod_{M+1}^{\infty} \frac{\mu_n}{\mu_n + \lambda_n} = V_M. \tag{8}$$

From (8), $h_M(t) \to 1$ as $M \to \infty$, so that (2) is verified. For (3), say $a \in A_N$, $b \in A_N$. For $M \geqq N$,

$$\sum_{c \in A_M} P(a, c, s)P(c, b, t)$$

$$= h_M(s)h_M(t) \sum_{\alpha_1, \cdots, \alpha_M} \prod_{n=1}^{M} R_n(\epsilon_n, \alpha_n, s)R_n(\alpha_n, \delta_n, t)$$

$$= h_M(s)h_M(t) \prod_{n=1}^{M} \left(\sum_{\alpha=0}^{1} R_n(\epsilon_n, \alpha, s)R_n(\alpha, \delta_n, t) \right)$$

$$= h_M(s)h_M(t) \prod_{n=1}^{M} R_n(\epsilon_n, \delta_n, s + t) \to P(a, b, s + t) \quad \text{as} \quad M \to \infty.$$

For (4), if $a \in A_N$ and $M \geqq N$,

$$P(a, a, t) \geqq \left(\prod_1^M R_n(\epsilon_n, \epsilon_n, t) \right) V_M,$$

so that

$$\liminf_{t \to 0} p(a,a,t) \geqq V_M.$$

Since this holds for all M and $V_M \to 1$ as $M \to \infty$, (4) is verified. Finally, since, for $a \in A_N$ and $M \geqq N$ we have, for all $k \geqq 1$

$$P(a, a, t) \leqq h_{M,k}(t) = \prod_{M+1}^{M+k} R_n(0, 0, t),$$

and since $P(a, a, 0) = h_{M,k}(0) = 1$,

$$P'(a, a, 0) \leqq h'_{M,k}(0) = -\sum_{M+1}^{M+k} \lambda_n,$$

so that (6) implies $P'(a, a, 0) = -\infty$.

REFERENCES

[1] R. L. Dobrudsin, "An example of a countable homogeneous Markov process all states of which are instantaneous", (Russian) *Teor. Veroyatnost. i Primenen.* 1(1956), 481–485.

[2] William Feller and Henry P. McKean, Jr., "A diffusion equivalent to a countable Markov chain," *Proceedings Nat. Acad. of Sci.*, Vol. 42, No. 6 (1956), pp. 351–354.

[3] P. Lévy, "Systems Markoviens et stationnaires. Cas denombrable," *Annales Sci. de l'Ecole Normale Superieure de Paris*, Vol. 68 (1951), pp. 327–381.

SPACINGS GENERATED BY MIXED SAMPLES

By Lionel Weiss[1]

Cornell University

1. Summary and introduction. Suppose $X(1, 1), X(1, 2), \cdots, X(1, n_1), X(2, 1), \cdots, X(2, n_2), \cdots, X(k, 1), \cdots, X(k, n_k)$ are independent chance variables, $X(i, j)$ having the probability density function $f_i(x)$, for $j = 1, \cdots, n_i$, $i = 1, \cdots, k$. We assume that for each i, $f_i(x)$ is bounded and has at most a finite number of discontinuities. We denote $n_1 + n_2 + \cdots + n_k$ by N, and we assume that n_i/N is equal to r_i, where r_i is a given positive number. Let $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_N$ denote the ordered values of the N observations

$$X(1, 1), \cdots, X(k, n_k).$$

Define W_i as $Y_{i+1} - Y_i$ for $i = 1, \cdots, N - 1$. For any given nonnegative t, let $R_N(t)$ denote the proportion of the values $W_1, \cdots, W_{N-1}$ which are greater than t/N. Let $S(t)$ denote

$$\int_{-\infty}^{\infty} (r_1f_1(x) + r_2f_2(x) + \cdots + r_kf_k(x)) \exp\{-t[r_1f_1(x) + \cdots + r_kf_k(x)]\}\, dx$$

and $V(N)$ denote $\sup_{t\geqq 0} | R_N(t) - S(t) |$. Then it is shown that $V(N)$ converges stochastically to zero as N increases. This is a generalization of [1], where k was equal to unity. The result is applied to find the asymptotic behavior of ranks in a k-sample problem.

2. Proof of the stochastic convergence of $V(N)$. As in [1], if it can be shown that $R_N(t)$ converges stochastically to $S(t)$ for each positive t, the convergence of $V(N)$ follows. Therefore we fix a positive value for t.

We define the chance variable $Z(i, j, N)$ to be equal to unity if no observations fall in the half-open interval $[(X(i, j), X(i, j) + t/N]$, and equal to zero otherwise. We denote $1/N\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z(i, j, N)$ by $K(N)$. Clearly,

$$K(N) = (1 - 1/N)R_N(t) + 1/N,$$

Received July 17, 1957; revised August 26, 1957.

[1] Research sponsored by the Office of Naval Research.

so our purpose is accomplished if we show that $K(N)$ converges stochastically to $S(t)$ as N increases.

We denote $\int_{-\infty}^{x} f_i(x)\,dx$ by $F_i(x)$.

$$E\{Z(i,j,N)\} = \int_{-\infty}^{\infty}\left[1 - F_i\left(x + \frac{t}{N}\right) + F_i(x)\right]^{n_i-1} \cdot \prod_{h\neq i}\left[1 - F_h\left(x + \frac{t}{N}\right) + F_h(x)\right]^{n_h} dF_i(x).$$

But with the exception of a finite number of points, $F_i(x + t/N) - F_i(x)$ can be written as $[f_i(x) + \epsilon_i(x, t/N)]t/N$, where $\epsilon_i(x, t/N)$ approaches zero as N increases, for each x. Since $f_i(x)$ is bounded $(i = 1, \cdots, k)$, it follows easily that $E\{Z(i, j, N)\}$ approaches

$$\int_{-\infty}^{\infty} \exp\{-t[r_1 f_1(x) + \cdots + r_k f_k(x)]\}\, dF_i(x)$$

as N increases. It follows immediately that $E\{K(N)\}$ approaches $S(t)$ as N increases.

Next we examine variance $\{K(N)\}$, which equals $N^{-2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}$ variance $\{Z(i, j, N)\} + 1/N^2 \sum\sum\sum\sum_{(i,j)\neq(g,h)} \text{cov}\,\{Z(i, j, N), Z(g, h, N)\}$. The first term in this last expression clearly approaches zero as N increases, since there are N uniformly bounded terms in the sum. We shall show that the second term also approaches zero by showing that the covariances approach zero uniformly. Since there are $N(N - 1)$ covariances, the factor $1/N^2$ guarantees the approach to zero. If $i \neq g$, $E\{Z(i, j, N)\cdot Z(g, h, N)\}$ is equal to

$$\iint_{|x-y|>\frac{t}{N}} \prod_{b\neq i,g}\left[1 - F_b\left(x + \frac{t}{N}\right) + F_b(x) - F_b\left(y + \frac{t}{N}\right) + F_b(y)\right]^{n_b}$$
$$\cdot\left[1 - F_i\left(x + \frac{t}{N}\right) + F_i(x) - F_i\left(y + \frac{t}{N}\right) + F_i(y)\right]^{n_i-1}$$
$$\cdot\left[1 - F_g\left(x + \frac{t}{N}\right) + F_g(x) - F_g\left(y + \frac{t}{N}\right) + F_g(y)\right]^{n_g-1} dF_i(x)\, dF_g(y).$$

By computations similar to those used on $E\{Z(i, j, N)\}$, it follows that

$$E\{Z(i, j, N)\cdot Z(g, h, N)\}$$

approaches

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-t[r_1 f_1(x) + \cdots + r_k f_k(x)]\}\cdot\exp\{-t[r_1 f_1(y) + \cdots + r_k f_k(y)]\} \cdot dF_i(x)\, dF_g(y)$$

and from this it follows that cov $\{Z(i, j, N), Z(g, h, N)\}$ approaches zero as N increases. In the same way, it follows that

$$\text{cov}\,\{Z(i, j, N), Z(i, h, N)\}$$

approaches zero ($j \neq h$). Thus variance$\{K(N)\}$ approaches zero as N increases, so $K(N)$ converges stochastically to $S(t)$, as does $R_N(t)$. Therefore we have shown that $V(N)$ converges stochastically to zero as N increases.

3. Application to ranks in k-samples. Define $T(i, j)$ as $F_1(X(i, j))$. Then $T(1, 1), \cdots, T(1, n_1)$ have unform distributions. Let $G_i(x)$ denote the resulting distribution function for $T(i, j)$. We assume that $G_i(x)$ allows a density function $g_i(x)$ (then $g_i(x)$ is zero outside the interval $[0, 1]$, is bounded, and has a finite number of discontinuities). Let $V_1 \leqq V_2 \leqq \cdots \leqq V_{N-n_1}$ denote the ordered values of $T(2, 1), \cdots, T(k, n_k)$, and let V_0 equal zero, V_{N-n_1+1} equal one. Let S_i denote the number of $T(1, j)$'s which lie in the interval

$$[V_{i-1}, V_i], \qquad i = 1, \cdots, N - n_1 + 1.$$

For each nonnegative integer r, let $Q_n(r)$ be the proportion of values among $S_1, \cdots, S_{N-n_1+1}$ which are equal to r. Define $g(y)$ as $\sum_{i=2}^{k} (r_i/(1 - r_1))g_i(y)$, and α as $(r_1/(1 - r_1))$. Define $Q(r)$ as

$$\alpha^r \int_0^1 \frac{g^2(y)}{[\alpha + g(y)]^{r+1}}\, dy.$$

Then it follows from the results above, using also the argument in [2], that $\sup_{r \geqq 0} | Q_N(r) - Q(r) |$ converges stochastically to zero as N increases. This can be used to show that certain tests of the hypothesis

$$F_1(x) = F_2(x) = \cdots = F_k(x)$$

are consistent. The discussion parallels that found in [2].

REFERENCES

[1] L. Weiss, "The stochastic convergence of a function of sample successive differences," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 532–536.

[2] J. R. Blum and L. Weiss, "Consistency of certain two-sample tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 242–246.

CORRECTION TO "AN EXTENSION OF THE KOLMOGOROV DISTRIBUTION"

By Jerome Blackman

Syracuse and Cornell Universities

1. Summary. It has been pointed out by J. H. B. Kemperman that an error in [1] invalidates the formulas arrived at in that paper. It is the purpose of this note to supply the correct formulas for the probabilities of Theorems 1 and 2. An Appendix by Professor Kemperman is included.

Received February 18, 1957; revised August 14, 1957.

2. Introduction. The major error in [1] lies in the mappings on pp. 516–517. Corrections can be made for this but unfortunately the resulting formulas are more complicated than before. A smaller error appears in the statement that $N(A_{2i}) = N(B_{2i})$, but this is easily corrected. The new formulas are so much more complicated that it has not seemed worthwhile to correct Corrolaries 1 and 2 which are hereby retracted. The corrected statements of the main results follow.

THEOREM 1. *Let* $x_1, x_2, \cdots x_n, x_1', x_2', \cdots x_{nk}'$ *be a sequence of* $n(k+1)$ *independent random variables with a common continuous distribution* $F(x)$. *Let* $F_n(x)$ *and* $G_{nk}(x)$ *be empiric distributions based on the first* n *and second* kn *random variables respectively. Then*

$$P(-y < G_{nk}(s) - F_n(s) < x \quad \text{for all} \quad s)$$
$$= 1 - \binom{(k+1)n}{n}^{-1} \sum_{i=1}^{\infty} \{N(A_{2i-1}) + N(B_{2i-1}) - N(A_{2i}) - N(B_{2i})\},$$

where the N *functions are given in* (1), (2), *and* (3).

THEOREM 2.

$$P(-y < F(s) - F_n(s) < x \quad \text{for all} \quad s)$$
$$= 1 - \sum_{i=1}^{\infty} \{\bar{N}(A_{2i-1}) + \bar{N}(B_{2i-1}) - \bar{N}(A_{2i}) - \bar{N}(B_{2i})\},$$

where the $\bar{N}$ *functions are given in* (5), (6), *and* (7).

3. Corrections. The point of departure from [1] is the middle of p. 516 where a formula for $N(\mathfrak{A}_0)$ is given. It is readily seen that upon dividing this equation by the total number of paths $\binom{(k+1)n}{n}$ one obtains Theorem 1 except for the analytical expressions for the N functions. We will also use the mapping of the A_i and B_i classes described at the bottom of p. 516, although the conclusions drawn there about the mapping are incorrect. The error is clear if we consider the image of a path from A_3 under the mapping. The image will be a path which starts from the origin, reaches $2\alpha + \beta$, and then on the return to 0 stops at least once at the point α. The class A_3 will be in 1:1 correspondence with the set of paths which starting at 0 reach $2\alpha + \beta$ and then on the return stop at least once at α. Because the steps to the left are of length k, not every path which reaches $2\alpha + \beta$ will, at some later step, stop at α. In Table 1 the images of the A_i and B_i under the mapping are given. The second column gives the points which the path must reach and the last column gives the points at which the path must stop, in order, after reaching the point described in column 2. In all cases the mapping is 1:1 between the class in the first column and the set of paths which reach the point indicated in the second column and subsequently stop in order at all the points indicated in the last column.

TABLE 1

A_{2i-1}	$i(\alpha+\beta)-\beta$	$(i-1)(\alpha+\beta)-\beta, \quad (i-2)(\alpha+\beta)-\beta, \cdots \alpha$
A_{2i}	$i(\alpha+\beta)$	$(i-1)(\alpha+\beta), \quad (i-2)(\alpha+\beta), \cdots (\alpha+\beta)$
B_{2i-1}	$i(\alpha+\beta)-\alpha$	$(i-1)(\alpha+\beta)-\alpha, \quad (i-2)(\alpha+\beta)-\alpha, \cdots \beta$
B_{2i}	$i(\alpha+\beta)$	$i(\alpha+\beta)-\beta, \quad (i-1)(\alpha+\beta)-\beta, \cdots \alpha$

As a preliminary step consider the number of ways a path consisting of i steps to the left and $ki-\alpha$ steps to the right can go from α to 0 without touching α after the first step. Let the number of these paths be $H_\alpha(i)$. While this number can be computed by elementary methods a more elegant formula has been obtained by Professor Kemperman, namely,

$$H_\alpha(i) = (k+1) \sum_{0 \leq r < \alpha/(k+1)} \frac{(-1)^r}{(i-r)(k+1)-1} \binom{(i-r)(k+1)-1}{i-r} \binom{\alpha-1-kr}{r} - \frac{\alpha}{(k+1)i-\alpha}\binom{(k+1)i-\alpha}{i}. \tag{1}$$

The proof of this is contained in the appendix.

The number of ways of going from 0 to α after exactly j steps to the left and $kj+\alpha$ steps to the right will be indicated by $J(\alpha, j)$ where

$$J(\alpha, j) = \binom{(k+1)j+\alpha}{j}. \tag{2}$$

Combining the results of Table 1 and the definitions of H and J we see that

$$\begin{aligned} N(A_{2i-1}) &= \sum_{j_1+\cdots+j_{i+1}=n} J(i(\alpha+\beta)-\beta, j_1) \prod_{k=2}^{i} H_{\alpha+\beta}(j_k) H_\alpha(j_{i+1}), \\ N(A_{2i}) &= \sum_{j_1+\cdots+j_{i+1}=n} J(i(\alpha+\beta), j_1) \prod_{k=2}^{i+1} H_{\alpha+\beta}(j_k), \\ N(B_{2i-1}) &= \sum_{j_1+\cdots+j_{i+1}=n} J(i(\alpha+\beta)-\alpha, j_1) \prod_{k=2}^{i} H_{\alpha+\beta}(j_k) H_\beta(j_{i+1}), \\ N(B_{2i}) &= \sum_{j_1+\cdots+j_{i+2}=n} J(i(\alpha+\beta), j_1) H_\beta(j_2) \prod_{k=3}^{i+1} H_{\alpha+\beta}(j_k) H_\alpha(j_{i+2}). \end{aligned} \tag{3}$$

This completes Theorem 1. The infinite series occurring in this theorem is really a finite series in view of $N(A_{2i-1}) = \cdots = N(B_{2i}) = 0$ For

$$i > nk/(\alpha+\beta).$$

To get Theorem 2 it is only necessary to take the limit as $k \to \infty$ in the various formulas given above. By Stirling's formula,

$$\binom{t(1+k)}{t} = \frac{t^t}{t!} O((1+k)^t) \quad \text{as} \quad k \to \infty. \tag{4}$$

Here and below we will use $a_k = 0(b_k)$ to mean $\lim_{k\to\infty} a_k/b_k = 1$.

Using (4) and a few more applications of Stirling's formula and remembering that $\alpha = -[-xkn]$ and $\beta = -[-ykn]$, we obtain

$$\begin{aligned} H_\alpha(i) &= \left\{ \sum_{0 \le r < xn} (-1)^r \frac{(i-r)^{i-r-1}(xn-r)^r}{(i-r)!r!} - xn(i-xn)^{i-1}/i! \right\} \\ &\qquad \cdot O((1+k)^i) = \bar{H}_x(i) O((1+k)^i) \end{aligned} \tag{5}$$

where the last equality defines $\bar{H}_x(i)$. Using (4) again

$$J(\alpha, j) = \frac{1}{j!}(j+xn)^j O((1+k)^j) = \bar{J}(x, j) O((1+k)^j) \tag{6}$$

and

$$\binom{(k+1)n}{n}^{-1} = \frac{n!}{n^n} O((1+k)^{-n}).$$

Combining these results and (3) the following equations are obtained:

$$\begin{aligned} &\lim_{k\to\infty} \binom{(k+1)n}{n}^{-1} N(A_{2i-1}) \\ &\qquad = \sum_{j_1+\cdots+j_{i+1}=n} \bar{J}(i(x+y)-y, j_1) \prod_{k=2}^{i} \bar{H}_{x+y}(j_k) \bar{H}(j_{i+1}) \\ &\qquad = \bar{N}(A_{2i-1}), \\ &\lim_{k\to\infty} \binom{(k+1)n}{n}^{-1} N(A_{2i}) \\ &\qquad = \sum_{j_1+\cdots+j_{i+1}=n} \bar{J}(i(x+y), j_1) \prod_{k=2}^{i+1} \bar{H}_{x+y}(j_k) \\ &\qquad = \bar{N}(A_{2i}), \\ &\lim_{k\to\infty} \binom{(k+1)n}{n}^{-1} N(B_{2i-1}) \\ &\qquad = \sum_{j_1+\cdots+j_{i+2}=n} \bar{J}(i(x+y)-x, j_1) \prod_{k=2}^{i} \bar{H}_{x+y}(j_k) \bar{H}_y(j_{i+1}) \\ &\qquad = \bar{N}(B_{2i-1}), \\ &\lim_{k\to\infty} \binom{(k+\ )n}{n}^{-1} N(B_{2i}) \\ &\qquad = \prod_{j_1+\cdots+j_{i+2}=n} \bar{J}(i(x+y), j_1) \bar{H}_y(j_2) \prod_{k=3}^{i+1} \bar{H}_{x+y}(j_k) \bar{H}(j_{i+2}) \\ &\qquad = \bar{N}(B_{2i}). \end{aligned} \tag{7}$$

This completes Theorem 2.

Attention should be drawn to a paper of Korolyuk [2] wherein the author gives different versions of the probabilities we have presented for the case $x = y$.

REFERENCES

[1] J. Blackman, "An extension of the Kolmogorov distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 513–520.

[2] V. S. Korolyuk, "On the difference of the empirical distribution of two independent samples," *Izv. Akad. Nauk. SSSR*, Vol. 19 (1955), pp. 81–96.

APPENDIX

By J. H. B. Kemperman

By a *path* of length n we shall mean an ordered sequence of $n + 1$ integers $(z_0, \cdots, z_n)$, such that

$$z_i - z_{i-1} \geqq -1 \qquad (i = 1, \cdots, n).$$

For each path $\pi_n = (z_0, \cdots, z_n)$, let

$$P(\pi_n) = \prod_{i=1}^{n} p(z_i - z_{i-1}),$$

(the weight or "probability" of π_n). Here, the $p_i = p(i)$, $(i = -1, 0, +1, \cdots)$, denote given (real or complex) numbers, $p(-1) \neq 0$. Finally, let

$$e_z(n) = \sum_{\pi_n}{}' p(\pi_n),$$

the summation being extended over all the paths $\pi_n = (z_0, \cdots, z_n)$ with $z_0 = 0$, $z_n = z$, $z_i \neq z (i = 0, 1, \cdots, n - 1)$.

Theorem. *For* $n = 1, 2, \cdots$,

$$e_z(n) = -zr_z(n)/n + \sum_{j=1}^{\infty} j(j+1)p_j \sum_{0<m\leqq +z} r_z(-m)r_{-j}(m+n-1)/(m+n-1). \tag{8}$$

Here, for arbitrary integers h and s, $r_h(s)$ *is defined as the coefficient of* w^{h+s} *in the formal development*

$$(p_{-1} + p_0 w + p_1 w^2 + \cdots)^s = \sum_h r_h(s) w^{h+s};$$

especially, $r_h(s) = 0$ *if* $h + s < 0$.

Proof. Let n and z be given integers, $n \geqq 1$. For any path $(z_0, \cdots, z_n)$ with $z_0 = 0$, $z_n = z$, we have

$$z_i - z_{i-1} = z - \sum_{\substack{\nu=1 \\ \nu \neq i}}^{n} (z_\nu - z_{\nu-1}) \leqq z + n - 1,$$

$(i = 1, \cdots, n-1)$, thus, $e_z(n)$ does not depend on the p_i with $i \geqq n + z$. Further, $r_h(s)$ does not depend on the p_i with $i \geqq h + s$, hence, the inner sum in (8) does not depend on the p_i with $i \geqq n + z$; moreover, the jth inner sum equals 0 when $j \geqq n + z$. Consequently, it suffices to prove the theorem for the special case that $p_i = 0$ for i sufficiently large.

In this case,

$$f(w) = \sum_{i=-1}^{\infty} p_i w^i$$

is analytic at each point $w \neq 0$. Further, for $|w|$ sufficiently small

$$f(w)^s = \sum_{-\infty}^{\infty} r_h(s) w^h, \tag{9}$$

hence, for $s \geqq 0$

$$r_h(s) = \sum_{\pi_s}{}' P(\pi_s),$$

summing over all the paths $\pi_s = (z_0, \cdots, z_s)$ with $z_0 = 0$, $z_s = h$. Observing that to each path $(z_0, \cdots, z_n)$ with $z_n = z$ there corresponds a unique integer m with $0 \leqq m \leqq n$, $z_i \neq z (i = 0, 1, \cdots, m-1)$, $z_m = z$, it follows that

$$r_z(n) = \sum_{m=0}^{n} e_z(m) r_0(n-m) \qquad (n = 0, 1, \cdots),$$

hence,

$$E_z = R_z / R_0, \tag{10}$$

where

$$R_h = \sum_{n=0}^{\infty} r_h(n) t^n, \qquad E_z = \sum_{n=0}^{\infty} e_z(n) t^n, \tag{11}$$

t denoting a sufficiently small parameter, $t \neq 0$.

Further, from (9), for each integer h,

$$R_h + \sum_{-h \leqq n < 0} r_h(n) t^n = \sum_{n=-h}^{\infty} r_h(n) t^n$$

$$= \sum_{n=-h}^{\infty} \frac{t^n}{2\pi\sqrt{-1}} \int_{|w|=R} f(w)^n w^{-h-1}\, dw = \frac{1}{2\pi\sqrt{-1}} \int_{|w|=R} \frac{(wf(w)t)^{-h}}{w - twf(w)}\, dw,$$

where R denotes a fixed positive number with $f(w) \neq 0$ for $0 < |w| \leqq R$. Here, from $p(-1) \neq 0$, the integrand is regular at $w = 0$. Moreover, for $t \neq 0$, $|t|$ sufficiently small, the equation $f(\xi) = t^{-1}$ has a unique solution satisfying $0 < |\xi| < R$. Thus,

$$R_h = (-t\xi f'(\xi))^{-1} \xi^{-h} - \sum_{0 < m \leqq h} r_h(-m) t^{-m}.$$

Finally, (10) and

$$\xi f'(\xi) = \xi f'(\xi) + f(\xi) - t^{-1} = -t^{-1} + \sum_{j=0}^{\infty} (j+1)p_j\xi^j$$

imply

$$E_z = \xi^{-z} + (-1 + t\sum_{j=0}^{\infty} (j+1)p_j\xi^j) \sum_{0<m\leq z} r_z(-m)t^{-m}.$$

In view of (11), it suffices to prove that, for each integer h and $|t|$ sufficiently small, $t \neq 0$,

$$\xi^{-h} = -h \sum_{\substack{m=-h \\ m\neq 0}}^{\infty} r_h(m)t^m/m + c_h ,$$

where c_h denotes a constant. Now, for $|t|, |\xi|$ small, the mapping $t \to \xi$ defined by $f(\xi) = t^{-1}$ is a 1:1 analytic transformation. Hence, integrating along a small positively oriented circle about 0, we have, for $m \neq 0$,

$$\int \xi^{-h}t^{-m-1}\, dt = -\int \xi^{-h}\, d(f(\xi)^m/m) = -\frac{h}{m}\int f(\xi)^m\xi^{-h-1}\, d\xi = -2\pi\sqrt{-1}\,\frac{h}{m}\, r_h(m).$$

REMARK. Results and methods analogous to the above may be found in the paper "The passage problem for a stationary Markov chain" by J. H. B. Kemperman, to appear in these Annals.

Let k be a fixed positive integer and choose $p(-1) = p(k) = 1$, $p(i) = 0$ for $i \neq -1, k$. Then $e_n(z)$ is equal to the number of sequences $(z_0, \cdots, z_n)$ with $z_i - z_{i-1} = -1$ or $+k$

$$(i = 1, \cdots, n), \quad z_0 = 0, \quad z_n = z, \quad z_i \neq z \qquad (i = 0, 1, \cdots, n-1).$$

Further, $H(i)$ is equal to the number of sequences $(z_n, z_{n-1}, \cdots, z_0)$ with

$$n = -\alpha + i(k+1) \geq 1, \qquad z_i - z_{i-1} = -1 \text{ or } k$$

$(i = 1, \cdots, n)$, $z_n = \alpha$, $z_0 = 0$, $z_i \neq \alpha (i = 0, \cdots, n-1)$. Hence,

$$H_\alpha(i) = e_\alpha(-\alpha + i(k+1))$$

and the above Theorem yields

$$H_\alpha(i) = -\alpha r_\alpha(n)/n + k(k+1) \sum_{0<m\leq\alpha} r_\alpha(-m)r_{-k}(m+n-1)/(m+n-1),$$

where $n = -\alpha + i(k+1)$. Noting that $r_h(s)$ is equal to the coefficient of w^{k+s} in the expansion of $(1 + w^{k+1})^s$ about 0, formula (1) easily follows.

ABSTRACTS

(Abstracts of papers presented at the Los Angeles Meeting of the Institute, December 27–28, 1957)

1. **Non-parametric Multiple-Decision Procedures for Selecting That one of K Populations Which has the Highest Probability of Yielding the Largest Observation.** (Preliminary Report) ROBERT BECHHOFER, Cornell University AND MILTON SOBEL, Bell Telephone Laboratories. (By title)

Let X_i be chance variables with density function $f_i(x)$, and let

$$p_i = \text{Prob}\,\{X_i > \max_{j \neq i} X_j\}\,(i = 1, \cdots, k).$$

Then $\sum_{i=1}^{k} p_i = 1$. Let $p_{[1]} \leqq \cdots \leqq p_{[k]}$ denote the ranked p_i. Let θ^*, P^* $(1 < \theta^* < \infty$, $1/k < P^* < 1)$ be specified constants. The goal is to select the population associated with $p_{[k]}$; the procedure must guarantee, (*) Prob {Correct Selection | $p_{[k]} \geqq \theta^* p_{[k-1]}\} \geqq P^*$. Procedure: "At the mth stage take the vector-observation $\mathbf{x}_m = (x_{1m}, \cdots, x_{km})$ where the x_{ij} $(j = 1, 2, \cdots)$ are independent observations from the ith population. Consider $\mathbf{y}_m = (y_{1m}, \cdots, y_{km})$ which is obtained by replacing the largest component of $\mathbf{x}_m$ by unity, and all other components by zero. Then $\mathbf{y}_m$ is an observation from a multinomial distribution with probability p_i associated with the ith component $(i = 1, 2, \cdots, k)$. (*) now can be guaranteed by continuing with procedures already proposed, e.g., these *Annals*, Vol. 27, p. 861. If $f_i(x) = g\{(x - \mu_i)/\theta\}$ $(i = 1, 2, \cdots, k)$, then the procedure can be used for selecting the population associated with the largest μ_i for any θ, known or unknown. Similar non-parametric procedures in which pairs of observations are taken from each population at each stage of experimentation, and which employ the range of each pair can be used for selecting that one of k populations which has the highest probability of yielding the largest sample range. If $f_i(x) = h\{(x - \mu_i)/\theta_i\}$ $(i = 1, 2, \cdots, k)$, then these latter procedures can be used for selecting the population associated with the largest θ_i for any set of μ_i, known or unknown. (Research supported in part by the U. S. Air Force through the Air Force Office of Scientific Research, ARDC, Contract No. AF 18(600)-331.) (Received September 25, 1957.)

2. **The Asymptotic Efficiency of Friedman's χ_r^2-test.** PH. VAN ELTEREN, Mathematical Centre, Amsterdam. (By title)

Let $F(x)$ be a continuous cdf with density function $f(x) = F'(x)$ and let

$$x_{\mu\nu}\ (\mu = 1, 2, \cdots, m;\ \nu = 1, 2, \cdots, n)$$

be a chance variable with distribution $F_{\mu\nu}(x) = F(x + \theta_\nu + \eta_\mu)$. It is assumed for convenience, that $\sum_\nu \theta_\nu = 0$. Friedman (1937) has constructed the χ_r^2-test for the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_n = 0$ (*J. Amer. Stat. Assn.*, Vol. 32, pp. 675–699). For alternatives $\theta_\nu = \theta_{\nu m} = \delta_\nu/\sqrt{m}$, where the δ_ν are given constants satisfying $\sum_\nu \delta_\nu = 0$, the asymptotic relative efficiency for $m \to \infty$ in the sense of Pitman of Friedman's test with respect to the corresponding 2-way-analysis of variance test is found to be $e_n = 12n(n + 1)^{-1}[\sigma \int f^2(x)\, dx]^2$, where σ^2 is the variance associated with $F(x)$. If $f(x)$ is normal, e_n reduces to $e_n = 3n/\pi(n + 1)$. (Received August 19, 1957.)

3. **Experiments With Mixtures.** HENRY SCHEFFÉ, University of California.

Experiments with mixtures of q components are considered, whose purpose is the empirical prediction of the response to any mixture of the components, when the response depends on the proportions $x_1, x_2, \cdots, x_q$ of the components present but not on the total amount. The factor space is then the $(q - 1)$-dimensional simplex where $x_1 + \cdots + x_q = 1$, $x_i \geq 0$. An experimental design called the *simplex lattice* and some modifications are treated; in the simplex lattice $x_i = 0, 1/m, 2/m, \cdots, 1$ for $i = 1, \cdots, q$ and some positive integer m, and the responses of all mixtures possible with these proportions are observed. The usual resolution of the response into general mean, main effects, and interactions does not seem possible, and so polynomial regression is employed. The problem of fitting an nth degree polynomial in $x_1, \cdots, x_q$ to the response is complicated by the fact that different polynomials give the same function on the simplex. Useful canonical forms are developed for $n \leq 3$. The coefficients in these forms are interpreted as various kinds of synergisms. The analysis of experiments with these designs leads to classes of polynomials orthogonal on the lattices. The paper will appear in *J. Royal Stat. Soc., Series B.* (Received October 25, 1957.)

4. Least-Squares Estimation when Residuals are Correlated. M. M. Siddiqui, University of North Carolina.

Let y_j, $j = 1, \cdots, N$ be observations on a variate and let $y_j = \sum_{i=1}^{p} \beta_i x_{ij} + \Delta_j$, $j = 1, 2, \cdots, N$, where x_{ij} are non-stochastic, and $\Delta' = (\Delta_1, \cdots, \Delta_N)$ is a $N(0, \sigma^2 P)$ vector, where 0 is a zero vector and P is an $N \times N$ correlation matrix. Using the usual least-squares estimates, b_i, of β_i which are obtained by minimizing $\sum \Delta_j^2$, and s^2 of σ^2, the covariance matrix of b_i is obtained for general P and bounds are set on these covariances by first obtaining the maxima and minima of a quadratic and a bilinear form $u'Au$ and $u'Av$ where u and v are $N \times 1$ vectors and A is an $N \times N$ real symmetric matrix under the conditions $u'u = v'v = 1$, $u'v = 0$. (Received October 31, 1957.)

5. A Property of Additively Closed Families of Distributions. Edwin L. Crow, Boulder Laboratories, National Bureau of Standards.

Consider a one-parameter additively closed family of univariate cumulative distribution functions $F(x; \lambda)$ (H. Teicher, *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778). Let three cumulants with orders in arithmetic progression exist and be non-zero. If all three orders are even, or if the first order is odd, it is also required that $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$. Consider linear combinations, with real, non-zero coefficients, of a finite number of independent variables with distributions in the family. It is proved that the only such linear combinations whose distributions are also in the family are those with coefficients unity. The additively closed families having this property may be called *strictly additively closed*. It can be shown that (one-parameter) additively closed stable families of distributions (normal and Cauchy in particular) with characteristic functions continuous in λ are not strictly additively closed, while Poisson, generalized Poisson, binomial, and gamma families are strictly additively closed. (Received October 31, 1957.)

6. Determining Sample Size for a Specified Width Confidence Interval. Franklin A. Graybill, Oklahoma State University.

If an experimenter decides to use a confidence interval to locate a parameter, he is concerned with at least two things: (1) Does the interval contain the parameter? (2) How wide is the interval? In general the answer to these questions cannot be given with absolute certainty, but must be given with a probability statement. The problem the experimenter then faces is: The determination of n, the sample size, such that (A) the probability will be equal to α that the confidence interval contains the parameter, and (B) the probability will be equal to β^2 that the width of the confidence interval will be less than d units (where α,

β^2, and d are specified). To solve this problem will generally require two things: (1) The form of the frequency function from which the sample of size n is to be selected; (2) Some previous information on the unknown parameters in the frequency function. This suggests that the sample be taken in two steps; the first sample will be used to determine the number of observations n to be taken in the second sample so that (A) and (B) will be satisfied. For a confidence interval on the mean of a normal population with unknown variance this problem has been solved by Stein for $\beta^2 = 1$. In this paper a theorem is proved which gives a method for determining n so that (A) and (B) will be satisfied. The theorem holds for parameters in the normal distribution and other distributions as well. (Received October 30, 1957.)

7. Nonparametric Estimation of Sample Percentage Point Standard Deviation. John E. Walsh, Lockheed Aircraft Corporation.

The available data consists of a random sample $x(1) < \cdots < x(n)$ from a reasonable well-behaved continuous statistical population. The problem is to estimate the standard deviation of a specified $x(r)$ that is not in the tails of the sample. The estimates examined are of the form $a[x(r + i) - x(r - i)]$ and the explicit problem consists of determining suitable values for a and i. The solution $a = (1/2)(n + 1)^{-3/10}\{[r/(n + 1)][1 - r/(n + 1)]\}^{1/2}$ and $i \doteq (n + 1)^{4/5}$ appears to be satisfactory. Then the expected value of the estimate equals the standard deviation of $x(r)$ plus $O(n^{-9/10})$; also the standard deviation of this estimate is $O(n^{-9/10})$. That is, the fixed and random errors for this point estimate are of the same order of magnitude with respect to n. Solutions can be obtained which decrease the order of one of these types of error. However, these solutions increase the order of the other type of error, so that the over-all error magnitude exceeds $O(n^{-9/10})$. (Received November 7, 1957.)

8. On the Structure of Distribution-Free Statistics. C. B. Bell, Xavier University of Louisiana and Stanford University.

Let $X_1, \cdots, X_n$ be a sample of a one-dimensional random variable X which has continuous cpf F. It has been observed that the distribution-free statistics commonly appearing in the literature can be written in the form $\Phi[F(X_1), \cdots, F(X_n)]$, where Φ is a measurable symmetric function defined on the unit cube. Such statistics are said to have structure (d). In establishing that having structure (d) is equivalent to being symmetric and strongly distribution-free for properly closed, symmetrically complete classes of cpf's, this paper extends a result of Birnbaum and Rubin while employing different methodology. These results interest a statistician because (1) they indicate that one should construct a statistic of structure (d) whenever one wishes to design a distribution-free statistic; and (2) they guarantee that each symmetric, strongly distribution-free statistic is of structure (d), and, hence, that the value of its cpf at any point is the volume of a polyhedral region in the unit cube. Under such circumstances the work of numerous statisticians indicate that it should be possible to evaluate the cpf explicitly; reduce it to recursion formulae; tabulate it with high-speed computers; or evaluate its limiting distribution. (Received November 7, 1957.)

9. On the Supremum of the Poisson Process. Ronald Pyke, Stanford University.

Let $\{X(t); t \geqq 0\}$ be a Poisson process (with shift) for which $\log E(e^{iwX(t)}) = -itw\alpha + \lambda t(e^{iw} - 1)$, $w \in R_1$, $\alpha, \lambda > 0$. Define $\sigma(x, T) = Pr\left\{\sup_{0 < t \leqq T} X(t) \leqq x\right\}$. Let $X_1, X_2, \cdots X_n$ be the ordered random variables of n independent and uniform $-$ (0, 1) random variables. The distribution function, $Pr\left\{\max_{1 \leqq i \leqq n} (ai - X_i) \leqq x\right\}$, is obtained for all $a, x \in R_1$. For $a = 1/n$, this reduces to the distribution function of D_n^+ (cf. Birnbaum and Tingey, A.M.S., Vol. 22). Utilizing this result, $\sigma(x, T)$ is obtained explicitly. Applications of these expressions to queueing theory and distribution-free statistics are given.

10. On the Distributions of Various Sums of Squares in an Analysis of Variance Table for Different Classifications With Correlated and Non-homogeneous Errors. B. R. BHAT, Karnatak University. (Preliminary Report) (By Title)

The distributions of various sums of squares in an analysis of variance table for two way classification have been obtained by Box (Ann. Math. Stat., Vol. 25, pp. 484–498) under the assumption that the vectors $X_{.j}$ for $j = 1, 2, \cdots q$ are independent vector observations from a p-variate normal population with mean μ and covariance matrix Σ. The vector $X_{.j}$ for each jth level of the factor B denotes p observations corresponding to the p levels of the other factor A. This paper gives the distributions of the various sums of squares for any n-way classification under similar normality assumptions. It is noted that these distributions, in general, follow a simple pattern and so is their mutual dependence. For $n = 3$, if we have a third factor C at r levels in addition to the above factors A and B and if we assume that $X_{..k}$ for $k = 1, 2, \cdots r$ are independent vector observations from a pq-variate normal population, then, according to the general pattern the first set consists of distributions of the sums of squares for the correction term, main effects A and B and their interaction $A\ B$. The second set (the only remaining set) consists of the distributions of the sums of squares for the remaining main factor C and its interactions with the effects in the first set. Any two distributions, not belonging to the same set are independent, whereas, the distributions in the same set are mutually dependent. (Received May 10, 1957)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

During 1957–58 T. W. Anderson will be a Fellow at the Center for Advanced Study in the Behavioral Sciences in Stanford, California.

Dr. Robert S. Aries is now Chairman of the Board of Aries Associates, whose offices for general consultation were recently transferred to 77 South Street, Stamford, Connecticut.

John Bailey has recently joined the staff of the Waltham Laboratories of Sylvania Electric Products, Inc., as an Engineer in their Applied Engineering Department.

Colin R. Blyth, on leave from the University of Illinois, will be at Stanford University for the academic year of 1957–58.

John V. Breakwell has taken a position as Staff Scientist with the Lockheed Missile Systems Division in Palo Alto.

D. M. Brown is now studying for the degree of Ph.D. in statistics at Princeton University on a RAND Corporation Fellowship.

Charles R. Carr has joined the research staff of The RAND Corporation at Santa Monica, California.

Richard L. Carter has received his Ph.D. degree in statistics from the University of North Carolina and has been appointed Associate Professor of Industrial Engineering at the Illinois Institute of Technology.

Jonas M. Dalton completed his work for his Master's degree in statistics at Virginia Polytechnic Institute in June, 1957. He is now employed at the Bell Telephone Laboratories, Murray Hill, New Jersey.

Morris H. De Groot has been appointed Assistant Professor in the Department of Mathematics at Carnegie Institute of Technology.

R. F. Drenick has joined the Bell Telephone Laboratories as a member of its technical staff.

Joseph Dubay is now an instructor in the Mathematics Department of the University of Oregon.

Professor Benjamin Epstein of Wayne State University is on leave at the Department of Statistics, Stanford University.

L. A. Gardner, Jr., has resigned his position as research scientist at Columbia University's Hudson Laboratories and is now employed as staff mathematician at M.I.T. Lincoln Laboratory.

John J. Gart is now a graduate fellow at the Oak Ridge Institute of Nuclear Studies, continuing work there toward a Ph.D. in statistics from V.P.I.

David W. Gaylor has resigned from the Nuclear Aircraft Research Facility, Convair, Ft. Worth, to work toward a Ph.D. in experimental statistics at North Carolina State College.

R. Gnanadesikan is now working with the Statistics Group at the Proctor and Gamble Company at Cincinnati, Ohio.

William A. Golomski, formerly Assistant Professor of Mathematics at Marquette University, is now in charge of operations research for Oscar Mayer and Company, Inc., Madison, Wisconsin.

Roe Goodman has gone from Santiago, Chile, where he was F.A.O. agricultural statistician, to Karachi, Pakistan, where he is now sampling statistician under the I.C.A. program of the U. S. Government in that country.

Ulf Grenander has accepted an appointment as Professor of Mathematical Statistics at Brown University.

Irwin Guttman is on leave of absence from the University of Alberta and will spend the academic year of 1957–58 as a Research Associate in the Department of Mathematics, Statistical Section, of Princeton University.

Following completion of assignment for Remington Rand Internation (installation of UNIVAC I at European Computing Center, Frankfurt/Main. Germany), Dr. Carl Hammer has accepted a similar position with Sylvania Electric Products, Inc., at their Waltham Laboratories.

Gordon M. Harrington has left his position as consultant in research, Connecticut State Department of Education, to become Associate Professor of Psychology and Department Chairman at Wilmington College, Wilmington, Ohio.

Dr. Theodore W. Horner, formerly of the Statistical Laboratory, Iowa State College, is now an Operations Research Analyst with General Mills, Inc., Minneapolis, Minnesota.

Patricia A. Inman received her M. A. in mathematics at U.C.L.A. in August, 1957, and has accepted a position as Computing Engineer at Atomics International, Canoga Park, California.

James Edward Jackson has taken a leave of absence from the Eastman Kodak Company to work on a doctorate at V.P.I. in Blacksburg, Virginia.

Trinidad J. Jaramillo is now working with System Research, University of Chicago, as Senior Research Engineer.

Peter W. M. John has accepted a position as research statistician with the California Research Corporation at Richmond, California.

Andre G. Laurent, formerly with the Department of Statistics of Michigan State University, has accepted an appointment as Associate Professor in the Department of Mathematics of Wayne State University.

J. Walter Lynch has moved from Huntsville, Alabama, to 101-L Rodman Road, Aberdeen, Maryland.

Albert Madansky has been employed by the Mathematics Division, The RAND Corporation, Santa Monica, California.

B. Mandelbrot, formerly of the University of Geneva, has accepted an appointment in the University of Lille.

Mr. C. L. Marcus has returned to the University of Illinois to complete his studies for a Ph.D. in statistics. He also holds an Assistantship in the Mathematics Department and a Fellowship from Armour Research.

Dr. H. P. Mulholland has been appointed to a Senior Lectureship in Mathematics in the University of Exeter.

Peter Newman has taken a post as Lecturer in Economics University College of the West Indies, Mona, St. Andrew, Jamaica, B.W.I.

Dr. Bernard Ostle, formerly Professor of Mathematics and Director of the Statistical Laboratory at Montana State College, is now with the Reliability Department of Sandia Corporation, Albuquerque, New Mexico.

Dr. Raymond P. Peterson has accepted a position as Mathematical Statistician with the Research Department, Matson Navigation Company, San Francisco, California.

Paul H. Randolph has resigned his position as Assistant Professor of Industrial Engineering at Illinois Institute of Technology to accept a position as Associate Professor of Industrial Engineering at Purdue University.

Dr. George J. Resnikoff, formerly Research Associate at the Applied Mathematics and Statistics Laboratory, Stanford University, has joined the Industrial Engineering Department of the Illinois Institute of Technology as Associate Professor.

Robert H. Riffenburgh received his Ph.D. degree in statistics at the Virginia Polytechnic Institute and is now Assistant Professor of Mathematics at the University of Hawaii, Honolulu.

Joseph Rosenbaum has resigned his position as Associate Mathematician, Systems Development Division, The RAND Corporation, and is presently employed as statistician, Broadview Research Corporation, Burlingame, California.

Dr. Jagdish S. Rustagi is an Assistant Professor in the Department of Statistics at Michigan State University, East Lansing, Michigan, for the current academic year.

Melvin E. Salveson has formed the Council for Advanced Management together with Herbert Holt, M.D., and is offering services for research in management and management education.

Dr. M. M. Siddiqui, who obtained his doctorate degree from the University of North Carolina in June, 1957, is now working for a temporary period with the Boulder Laboratories of the National Bureau of Standards in the capacity of Mathematical Statistician.

Professor Jack Wilber has returned to Roosevelt University after spending four months as Consultant to the Operations Analysis Office at the Air Force Missile Test Center.

Morris Skibinsky has returned to the Statistical Laboratory at Purdue after a year's leave of absence at Michigan State University.

James H. Stapleton received his Ph.D. degree in mathematical statistics from Purdue University in June, 1957, and is now a statistical consultant in the statistical methods section of General Electric's General Engineering Laboratory in Schenectady, New York.

Daniel Teichroew has joined the Graduate School of Business at Stanford University as Associate Professor of Management.

James E. Thompson has returned to his job as mathematician with the Defense Department, having completed a year of graduate study with the Department of Statistics at Stanford University on a Defense Department fellowship.

W. A. Thompson, Jr., has left the U. S. Air Defense Board and has accepted an academic position at the University of Delaware.

Dr. Fred H. Tingey, mathematician, has been appointed by TECHNICAL OPERATIONS, INCORPORATED, as Assistant Chairman of Experiment Planning and Execution, Dr. Ian W. Tervet, Director of the research and development firm's West Coast office, announced today. Dr. Tingey received his masters and doctoral degrees in Mathematics and Mathematical Statistics at the University of Washington. He graduated from Utah State College in 1947. In his new position, Dr. Tingey will plan and direct field experiments conducted by TECHNICAL OPERATIONS in conjunction with the Combat Development Experimentation Center (CDEC) at Fort Ord, California.

John A. Tischendorf, having completed two years of active duty with the Commissioned Corps, U. S. Public Health Service, has joined the staff of the Allentown Laboratory of Bell Telephone Laboratories, Inc.

Joseph B. Tysver, formerly an Associate Research Engineer at the University of Michigan, received his Ph. D. degree from that University in June, 1957.

and has accepted a position as Research Specialist in the Pilotless Aircraft Division of Boeing Airplane Company, Seattle, Washington.

Dr. John S. White has accepted a position as statistician with the Aero Division of Minneapolis-Honeywell Regulator Company.

Robert A. Wijsman is now Assistant Professor in the statistics group of the Department of Mathematics at the University of Illinois.

New Members

The following persons have been elected to membership in The Institute

August 7, 1957, to November 1, 1957

Abraham, T. C., M. Sc. (Karnatak Univ., India), Teaching Fellow, Boston University Graduate School, Boston University, Boston 15, Massachusetts; *627 Commonwealth Ave., Boston 15, Mass.*

Albert, Arthur E., M. S. (Stanford Univ.), Student, Department of Statistics, Stanford University, Stanford, California; *20 Russell Ave., Portola Valley, Calif.*

Ali, Asghar, M. A. (Univ. of North Carolina), Lecturer in Statistics, *Institute of Statistics, University of the Panjal, Lahore, Pakistan.*

Anglin, Ernie LaRue, B. S. (Univ. of Georgia), Student, *Department of Mathematics, University of Georgia, Athens, Georgia.*

Beatty, Richard L., M. S. (Univ. of Colorado), *Instructor in Statistics, University of Wyoming, Laramie, Wyoming.*

Bland, Richard P., B. S. (Univ. of Oklahoma), Student, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina; *1 Audley Lane, Chapel Hill, N.C.*

Bobotek, Henry G., M. A. (Univ. of Illinois), Research Associate, *Control Systems Laboratory, University of Illinois, Urbana, Illinois.*

Bogdanoff, J. L., Ph. D. (Columbia Univ.), *Professor of Engineering Sciences, Purdue University, Lafayette, Indiana.*

Calhoun, David W., B. A. (Yale Univ.), Biometrician, G. D. Searle and Co., P. O. Box 5110, Chicago 80, Illinois; *820 Hamlin St., Evanston, Ill.*

Caspers, James W., M. S. in E. E. (Univ. of Washington), Head, Applied Theoretical Studies Group, U. S. Navy Electronics Lab., San Diego 52, California; *5014 August St., San Diego 10, Calif.*

Champernowne, D. G., M. A. (Cambridge Univ.), Professor in Statistics and Fellow of Nuffield College, *Oxford University, Nuffield College, Oxford, England.*

Chapman, James W., M. S. (Univ. of Minnesota), Research Assistant, *Department of Soils, Institute of Agriculture, University of Minnesota, St. Paul 1, Minnesota.*

Clarke, Geoffrey M., M. A. (Oxon), Statistician, Department of Agriculture and Horticulture, National Fruit and Cider Institute, University of Bristol; *University of Bristol, Research Station, Long Ashton, Bristol, England.*

Cotton, John W., Ph. D. (Indiana Univ.), Assistant Professor of Psychology, Northwestern University, Evanston, Illinois; *Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois.*

Cox, Constance E., M. S. (Iowa State College), *Head, Biometrics Section, Food and Drug Directorate, Department of National Health and Welfare, Tunney's Pasture, Ottawa, Ontario, Canada.*

Dear, Robert E., Ph. D. (Univ. of Washington), Research Associate, *Research Division, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.*

Engler, Jean, Ph. D. (Northwestern Univ.), National Science Foundation Postdoctoral

Fellow, *Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.*

Evans, Richard V., A. B. (Princeton Univ.), Student, Department of Industrial Engineering, Johns Hopkins University, Baltimore 18, Md., *Barroll Road, Baltimore 9, Md.*

Feldt, Leonard S., Ph. D. (State Univ. of Iowa), Assistant Professor, *College of Education, State University of Iowa, Iowa City, Iowa.*

Fields, Raymond I., M. A. (Univ. of Arizona), Student, Virginia Polytechnic Institute, Blacksburg, Virginia; *Speed Scientific School, University of Louisville, Louisville, Kentucky.*

Figge, Harry J., President, *Harry J. Figge and Associates, 800 Liberty Building, Des Moines 9, Iowa*

Flanagan, Richard M., B. A. (Univ. of Michigan), Senior Programmer, *Argonaut Insurance Group, 250 Middlefield Road, Menlo Park, California.*

Gordon, M. H., Ph. D. (Univ. of Tennessee), Assistant Director, Central Neuropsychiatric Research Unit, Veterans Administration Hospital, Perry Point, Maryland; *Box 546, Perry Point, Maryland.*

Govindarajulu, Z., M. A. (Univ. of Madras), Graduate Student, Statistics, and Teaching Assistant, *Biostatistics Division, University of Minnesota, Minneapolis 14, Minnesota.*

Green, Lloyd G., B. A. (Washington Missionary College), Mathematician, *Touche, Niven, Bailey and Smart, 1380 National Bank Building, Detroit 26, Michigan.*

Halton, John H., M. A. (Oxon), Research-Student in Faculty of Physical Sciences, *Balliol College, Oxford University, Oxford, England.*

Hancock, John V., B. S. (Memphis State Univ.), Research Assistant, *Department of Mathematics, University of Georgia, Athens, Georgia.*

Harrison, Gerald, Ph. D. (Calif. Institute of Technology), Mathematician, *The Teleregister Corp., 445 Fairfield Avenue, Stamford, Connecticut.*

Heinhold, Josef, Dr. rer. nat. (Technische Hochschule Munchen), Professor, *Institut fur Angewandte Mathematik, Technische Hochschule Munchen, Munchen 2, NW, ArcisstraBe 21, Germany.*

Hicks, Charles R., Ph. D. (Syracuse Univ.), Associate Professor of Mathematics and Research Associate in the Statistical Laboratory, *Statistical Laboratory, Engineering Administration Building, Purdue University, Lafayette, Indiana.*

Hoyland, Arnljot, Cand. real (Univ. of Oxlo), Research Assistant, Forsikringsteknish Seminar, University of Oslo, Blindern pr. Oslo, Norway; *Krokvolden 20, Stabekk pr. Oslo, Norway*

Iversen, Iver Andrew, B. A. (Univ. of Minnesota), Teaching Assistant, University of Minnesota, Minneapolis 14, Minnesota; *420 Fifth Street S.E., Minneapolis 14, Minnesota.*

Jacobsen, Fred M., Jr., Ph. D. (Iowa State College), Group Leader, Computer Programming and Mathematical Analysis, American Oil Co., Box 401, Texas City, Texas; *Box 1537, Texas City, Texas.*

Jones, Alfred Welwood, Ph. D. (Columbia Univ.), Systems Engineer, *Bell Telephone Laboratories, 463 West Street, New York 14, N. Y.*

Kakeshita, Shin'ichi, B. Sc. (Kyushu Univ.), Student, *Math. Inst., Fac. Sci., Kyushu University, Fukuoka, Japan.*

Kim, Dong Sie, B. S. (Seoul National Univ.), Assistant, Dept. of Mathematics, Seoul National University, Seoul, Korea; *11–44 Ka-heo-Dong, Chong-no-Ku, Seoul, Korea.*

Knapp, Leslie E., B. S. and B. A. (Upper Iowa Univ.), Student, Stanford University, Stanford, California; *1255 Tucson Avenue, Sunnyvale, California.*

Lamm, Richard A., M. A. (Hofstra College), Analytical Statistician, Chemical Corps R and D Command, Biological Warfare Laboratories, Fort Detrick, Frederick, Maryland; *75 Stewart Manor, Frederick, Maryland.*

Laubscher, N. F., M. Sc. (Potchefstroomse Universiteit vir C.H.O.), A. Research Officer, *South African Council for Scientific and Industrial Research, National Physical Research Laboratory, P. O. Box 395, Pretoria, South Africa.*

Lindeman, Richard H., M. S. (Univ. of Wisconsin), Research Fellow, Bureau of Institutional Research, University of Minnesota, Minneapolis 14, Minnesota; *4936 Penn Ave. South, Minneapolis 9, Minn.*

Lokki, Olli Kristian, Dr. phil. (Univ. of Helsinki), *Associated Professor, Institute of Technology, Helsinki, Finland.*

Lunneborg, Clifford E., B. S. (Univ. of Washington), Research Assistant, Division of Counseling and Testing Services, University of Washington, Seattle, Washington; *5218 16th Avenue N.E., Seattle 5, Washington.*

McGuire, Judson Ulery, Jr., Ph. D. (Iowa State College) Entomologist, Agricultural Research Service, U. S. Department of Agriculture, Washington, D. C.; Apartado 654, Camaguey, Cuba.

Mikami, Misao, D. Sc. (Kyushu Univ.), Professor of Industrial Statistics, *Seminar of Industrial Statistics, Faculty of Engineering, Kyushu University, Fukuoka, Japan.*

Mitra, S. S., M. S. (Univ. of Calcutta), Graduate Student and Teaching Assistant, *Department of Mathematics, University of Washington, Seattle 5, Washington.*

Miyasawa, Koichi, D. Sc. (Kyushu Univ.), Assistant Professor of Mathematical Statistics and Econometrics, *Faculty of Economics, Tokyo University, Tokyo, Japan.*

Pendergrass, R. N., M. A. (Univ. of Missouri), *Professor of Mathematics, Radford College, Radford, Virginia.*

Pike, M. C., B. S. (Witwatersrand Univ., South Africa), Student, Statistical Laboratory, University of Cambridge, *Trinity Hall, Cambridge, England.*

Redus, Faye, B. S. (Stephen F. Austin State College), Senior Analyst and Programmer, Sutherland Co., Suite 1112, First National Bank Building, Peoria, Illinois; *434 W. 21 Street, San Bernardino, California.*

Reed, James C., Ph. D. (Univ. of Chicago), *Director, Reading and Study Skills, Wayne State University, Detroit, Michigan.*

Rice, Philip L., B. S. (Principia College), Chief, Tropospheric Analysis Section, Radio Propagation Engineering Division, National Bureau of Standards, Boulder, Colorado; *1103 Pine Street, Boulder, Colorado.*

Rogerson, G. W., B. S. (Melbourne), Student, Melbourne University, Carlton, Melbourne N3, Victoria, Australia; *48 Drummond St., Carlton, Melbourne N3, Victoria, Australia.*

Romano, Albert, M. A. (Washington Univ.), Student and Research Assistant, *Dept. of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.*

Sagi, Philip C., Ph. D. (Univ. of Minnesota), Research Associate, *Office of Population Research, 5 Ivy Lane, Princeton, New Jersey.*

Schoderbek, Joseph J., M. S. (Carnegie Inst. of Tech.), Research Engineer, Missile Systems Division, Lockheed Aircraft Corp., Palo Alto, California; *543 Carla Court, Mountain View, Calif.*

Schwartz, A. J., B. S. (Wayne Univ.), Student, Wayne State University, Detroit, Michigan; *18478 Prest, Detroit, Michigan.*

Shaffer, Douglas H., Ph. D. (Carnegie Inst. of Technology), Mathematician, *Westinghouse Research, Pittsburgh 35, Pa.*

Sherman, Seymour, Ph. D. (Cornell Univ.), Professor, *Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia 4, Pennsylvania.*

Singh, Rajinder, M. A. (Panjale Univ., India), Graduate Assistant, *University of Illinois, Department of Mathematics, Urbana, Illinois.*

Spicer, Ira G., B. S. (Univ. of Minnesota), Development Engineer, Project Leader of Technical Analysis, Minneapolis-Honeywell, Ordnance Engineering, Hopkins, Minnesota; 1928 Emerson Ave. So., Apt. 1-D, *Minneapolis 5, Minnesota.*

Trammell, Carol D., B. S. (Carnegie Inst. of Tech.), Graduate Student and Teaching Assistant, *Department of Mathematics, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.*

Weiler, H., M. S. (N.S.W. Univ. of Technology), Research Officer, *CSIRO, Sheep Biology Laboratory, P. O. Box 144, Paramatta, N.S.W., Australia.*

Woods, W. Max, M. S. (Oregon Univ.), Student, Stanford University, Stanford, California; *1919 Manhattan Ave., Apt. 4, East Palo Alto, California.*

Youtcheff, John S., A. B. (Columbia Univ.), Operations Analyst, General Electric Company, Missile and Ordnance Systems Dept., Philadelphia, Pennsylvania; *Post Office Box 155, Berwyn, Pennsylvania.*

Congratulations to the Office of Naval Research

At the request of the Council, the President of the Institute of Mathematical Statistics has written a congratulatory letter to Admiral Bennett of the Office of Naval Research in connection with the tenth anniversary of the Office of Naval Research. The text of the letter follows.

Dear Admiral Bennett:

At the recent annual meeting of the Institute of Mathematical Statistics, the Council unanimously asked me to offer our congratulations and best wishes on the tenth anniversary of the establishment of the Office of Naval Research.

Through its support of senior investigators and graduate students and the consequent publication of many important technical papers and books, the Office of Naval Research has been contributing greatly to the advancement of fundamental research in mathematical statistics and probability theory. This contribution is especially important because it is being made during a period when these fields are showing themselves capable of particularly rapid growth.

The help you have given our profession is but one aspect of that program through which government and science work hand in hand to the benefit of each and to that of the nation as a whole, both in military and civilian pursuits. The Navy Department is outstanding in this respect, and it must be pleased and honored by the record of your Office.

Congratulations to you and your staff and best wishes for many more years of success in your efforts.

Respectfully yours,
Leonard J. Savage
President

Fifth Annual Southern Regional Graduate Summer Session in Statistics

The fifth Southern Regional Graduate Summer Session in Statistics will be held June 16 through July 26, 1958, at Oklahoma State University, Stillwater, Oklahoma. The summer sessions are rotated annually among the following institutions: Virginia Polytechnic Institute, Oklahoma State University, University of Florida and North Carolina State College.

The program may be entered at any session, and consecutive courses will be offered in successive summers. The summer work in statistics may be applied towards residence requirements at any one of the cooperating institutions, as well as certain other institutions, in partial fulfillment of residence requirements for graduate degrees. Each annual summer session lasts six weeks and the several courses offered carry three semester hours of graduate credit.

The summer sessions are designed to carry out a recommendation of the Southern Regional Education Board's Committee on Statistics, on which the four institutions initiating the program are represented.

The sessions will be of particular interest to (1) research and professional workers who want intensive instruction in basic statistical concepts and who wish to learn modern statistical methodology, (2) teachers of elementary statistics courses who want some formal training in modern statistics, (3) prospective candidates for graduate degrees in statistics, (4) graduate students in other fields who desire supporting work in statistics, and (5) professional statisticians who wish to keep informed of advanced specialized theory and methods.

The faculty for the 1958 Summer Session at Oklahoma State University will include the following visiting professors: H. O. Hartley, Statistical Laboratory, Iowa State College; Walter T. Federer, Biometrics Unit, Cornell University; John E. Freund, Department of Mathematics, Arizona State College; A. W. Wortham, Operations Research Department, Texas Instruments, Dallas, Texas.

The local staff includes: Carl E. Marshall (Ph.D., Iowa State), Franklin Graybill (Ph.D., Iowa State), Robert D. Morrison (Ph.D., North Carolina State, John Hamblen (Ph.D., Purdue), Roy Deal (Ph.D., University of Oklahoma), and John Hoffman (Ph.D., University of Oklahoma).

Of particular interest at this summer session will be the six weekly symposia covering six important areas in statistics. They are: Sampling Survey Designs, Experimental Designs, Non-parametric Statistics, Response Curves and Surfaces, Multiple Comparisons, and High Speed Computing. Discussants will be selected from major contributors to these areas. These invited speakers together with the outstanding summer school staff will cover the respective subjects from three points of view: applications, their mathematical bases, and the problems that lie on the frontier.

Inquiries should be addressed to Carl E. Marshall, Director, Statistical Laboratory, Oklahoma State University, Stillwater, Oklahoma.

List of International and Foreign Scientific and Technical Meetings

October 1, 1957 through 1960

(The following information was extracted from a list compiled by the Office of Scientific Information of the National Science Foundation.)

Date and Place	*Meeting, Sponsor and Subject*	*Address Queries to:*
Oct. 21, 1957 Lourenco Marques, Mozambique	2nd Inter-African Conference on STATISTICS, Inter-African Committee of Statistics	Secretariat, Commission for Technical Cooperation in Africa South of the Sahara, 43 Parliament Street, London, S. W. 1, England

Date and place	*Meeting, Sponsor and Subject*	*Address Queries to:*
Dec. 26, 1957–Jan. 4, 1958 Berkeley, California	Symposium on AXIOMATIC METHOD, with Special Reference to Geometry and Physics	Professor Alfred Tarski, Department of Mathematics, University of California, Berkeley 4, California
Apr. 9–13, 1958 Giessen, Germany	Society for APPLIED MATHEMATICS AND MECHANICS (GAMM), Annual Meeting	Professor Dr. Egon Ullrich, Mathematisches Institut der Justus Liebig-Hochschule, Johannesstrasse 1, (16) Giessen, Germany
Apr. 13–16, 1958 Giessen, Germany	German MATHEMATICS SOCIETY (DMV), Annual Meeting	Professor Dr. Egon Ullrich, Mathematisches Institut der Justus Liebig-Hochschule, Johannesstrasse 1, (16) Giessen, Germany
June 1958 Strasbourg, France	International Association for ANALOGY COMPUTATION, 2nd International Meeting and 1st General Assembly	Professor J. Hoffmann, Universite Libre, 50 Avenue Franklin D. Roosevelt, Brussels, Belgium
Aug. 11–13, 1958 St. Andrews, Scotland	International MATHEMATICAL Union, 3rd General Assembly	Mr. F. Smithies, Mathematical Institute, 16 Chambers Street, Edinburgh 1, Scotland
Aug. 14–21, 1958 Edinburgh, Scotland	11th International Congress of MATHEMATICIANS—Logic and foundations, algebra and theory of numbers; analysis; topology; geometry; probability and statistics; applied mathematics, mathematical physics and numerical analysis; and history and education	Mr. F. Smithies, Mathematical Institute, 16 Chambers Street, Edinburgh 1, Scotland
Sept. 3–10, 1958 Namur, Belgium	2nd International Congress for CYBERNETICS, Association Internationale de Cybernetique (ASBL)—Information—Automatism (Cybernetics applied to machinery)—Automation (Cybernetics used in organizing labor)—The economical and social consequences of Automation—Cybernetics and social sciences — Cybernetics and biology	International Association for Cybernetics, 13, rue Basse-Marcelle, Namur, Belgium
Sept. 1958 Brussels, Belgium	International STATISTICAL Institute, Special Session	Institut National de Statistique, 44, rue de Louvain, Brussels, Belgium
1958 Undecided Warsaw, Poland	International Symposium on NONHOMOGENEITY IN ELASTICITY AND PLASTICITY, International Union for Theoreti-	Dr. Hugh L. Dryden, President of Union, NACA, 1512 H Street, N. W., Washington, D. C.; or Professor F. N. van

Date and place	*Meeting, Sponsor and Subject*	*Address Queries to;*
	cal and Applied Mechanics (IUTAM)	den Dungen, Secretary of Union, 41 avenue de l'Arbalete, Boitsfont, Brussels, Belgium
1960 Undecided Stresa, Italy	10th International Congress of APPLIED MECHANICS, International Union of Theoretical and Applied Mechanics (IUTAM)	Dr. Hugh L. Dryden, President of Union, NACA, 1512 H Street, N. W., Washington, D. C.; or Professor F. N. van den Dungen, Secretary of Union, 41 avenue de l'Arbalete, Boitsfont, Brussels, Belgium

Royal Statistical Society Research and Industrial Applications Sections

The Research Section and the Industrial Applications Section of the Royal Statistical Society intend to hold a Conference at the University of St. Andrews, near Edinburgh, Scotland, from 22 August to 1 September inclusive. It will be devoted to Mathematical Statistics, with special reference to those branches of the subject which have application in industry.

It is proposed that there should be three morning sessions, consisting each of two or three pre-arranged lectures, and three early evening sessions (5.30 p.m. to 6.30 p.m.) each with one pre-arranged lecture.

The afternoons will be devoted to 'Splinter Groups' which will devote themselves to special aspects, and at which informal talks of some ten or fifteen minutes each can be given without prior arrangement.

Topics to be covered in the morning and evening sessions include aspects of the analysis of variance, non-parametric inference, stochastic aspects of linear and dynamic programming, and foundations of probability in statistics.

It is hoped that many of the mathematical statisticians who will be coming from abroad to attend the Edinburgh International Congress of Mathematicians, will choose to remain in Scotland for a further few days, and take the opportunity of meeting colleagues specially interested in their field. St. Andrews, besides having the famous Golf Course, is a small Scottish town of considerable character, and a very good centre for the exploration of the Eastern Highlands.

Accommodation (from 21 August to 2 September) will be provided within the hostels of the University of St. Andrews at a reasonably low cost, details to follow later. Anyone interested should write, marking the envelope 'ST. ANDREWS CONFERENCE' to Miss U. Croker, Royal Statistical Society, address as above.

University of Michigan Graduate School of Public Health Summer Program

The University of Michigan is the host institution for a cooperative program by the accredited Schools of Public Health of the United States during the summer of 1958.

The summer program for 1958 is designed to meet some of the educational and training needs of men and women engaged in work in health and health related agencies or those preparing themselves for such work. Courses are offered at three levels. The elementary level courses are intended for those who have acquired little or no background in statistical methodology. Intermediate courses present subject matter to extend and improve knowledge and skills of those persons who have acquired the elementary concepts and skills of statistical methodology. The advanced level courses are for those who have acquired considerable background in the theory and application of statistical concepts and procedures. A seminar open to all students includes the presentation of topics of current national interest related to health sciences and statistical methodology.

The faculty will consist of Helen Abbey, The Johns Hopkins University; William G. Cochran, Harvard University; Jerome Cornfield, The National Institutes of Health; Bernard Greenberg, University of North Carolina; F. M. Hemphill, University of Michigan; Leslie Kish, University of Michigan; Donovan Thompson, University of Pittsburgh; Colin White, Yale University.

If possible, completed applications and transcripts should reach Ann Arbor by June 1, 1958, for Michigan residents and May 1, for nonresidents. Requests for application forms should be addressed to the Director of the Summer Program in Public Health Statistics, School of Public Health, University of Michigan, Ann Arbor, Michigan.

A limited number of scholarships will be available to qualified students taking courses for credit. Inquiries concerning scholarships should be addressed to Dr. F. M. Hemphill, Director of the Summer Program in Public Health Statistics, School of Public Health, University of Michigan, Ann Arbor, Michigan.

IMS MEMBERS ATTENDING THE 1957 ANNUAL MEETING OF THE IMS

(This list was not received in time to be included with the report in the December, 1957 issue.)

Forman S. Acton, Frank B. Akutowicz, William R. Allen, Allan G. Anderson, R. L. Anderson, Virgil L. Anderson, William B. Anderson, Abdur Rahman Ansari, F. J. Anscombe, Kenneth J. Arnold, Samuel I. Askovitz, Ralph Hoyt Bacon, J. C. Bain, Theodore A. Bancroft, Rolf E. Bargmann, Glenn E. Bartsch, W. D. Baten, Grace E. Bates, Geoffrey Beall, Helen P. Beard, Robert Eric Bechofer, Charles Bernard Bell, Jr., Irving Belson, Andrew Angelo Benvenuto, Agnes P. Berger, Joseph Berkson, Gerald D. Berndt, Max A. Bershad, Reid A. T. Bhaucha, Charles A. Bicking, Patrick Paul Billingsley, Richard S. Bingham, Jr., Allan Birnbaum, David Blackwell, Herman Blasbalg, Chester I. Bliss, Julius R. Blum, Isadore Blumen, John B. Boddie, Derrill Joseph Bordelon, Raj C. Bose, Helen Bozivich, Ralph Allan, Bradley, A. E. Brandt, Leroy S. Brenna, Glenn W. Brier, Harold F. Bright, Samuel H. Brooks, Bernice Brown, Irwin D. J. Bross, Benjamin Buchbinder, Robert W. Burgess, Paul J. Burke, Irving W. Burr, Glenn L. Burrows, Lyle D. Calvin, Burton H. Camp, C. S. Callard, Mavis B. Carroll, Marvin F. Carter, Jack Chassan, Herman Chernoff, Victor Chew, Chin Long Chiang, John T. Chu, Joseph Louis Ciminera, Ira H. Cisin, Willard

H. Clatworthy, Andrew G. Clark, Charles William Clunies-Ross, William G. Cochran, Paul M. Cohen, William S. Connor, Clyde H. Coombs, Lewis C. Copeland, Richard G. Cornell, Jerome Cornfield, L. M. Court, Edwin L. Cox, Paul Charles Cox, Allen T. Craig, Elliot M. Cramer, Jean A. Crockett, Lee S. Crump, Edward Eugene Cureton, Joseph F. Daly, Cuthbert Daniel, Herbert Theodore David, Willis L. Davis, Read B. Dawson, Jr., W. Edwards Deming, Arthur P. Dempster, Cyrus Derman, Lucile Derrick, Earl Louis Diamond, John K. Diederichs, James L. Dolby, Tom G. Donnelly, Acheson J. Duncan, David B. Duncan, Paul R. Dunlap, Charles W. Dunnett, David Durand, Arthur Morlan Dutton, Meyer Dwass, Albert Ross, Eckler, Jr., Sylvain Ehrenfeld, Churchell Eisenhart, Harry Eisenpress, Salah A. Elmaghraby, Lila R. Elveback, Daniel R. Embody, Walter T. Federer, A. V. Fend, Robert Ferber, George Emery Ferris, William B. Fetters, Donald Fraser, David Frazier, Spencer M. Free, Spencer Michael Free, Jr., Agnes M. Galligan, Donald A. Gardiner, Werner Gautschi, Charles E. Gates, Donald Paul Gaver, Seymour Geisser, Lincoln J. Gerende, George William Gershefski, B. C. Getchell, Walter M. Gilbert, Dorothy Morrow Gilford, Leon Gilford, Harold Glazer, William A. Glenn, Ramanathan Gnanadesikan, Leo A. Goodman, Mina H. Gourary, Bernard G. Greenberg, Samuel W. Greenhouse, Joseph Arthur Greenwood, Frank E. Grubbs, Lee Gunlogson, John Gurland, Donald Guthrie, Irwin Guttman, Robert John Hader, Max Halperin, James F. Hannan, Morris Howard Hansen, Robert H. Hanson, Bernard Harris, T. E. Harris, Boyd Harshbarger, H. Leon Harter, Herman O. Hartley, William C. Healy, Jr., Paul Heit, F. M. Hemphill, G. Ronald Herd, Irene Hess, Clifford Hildreth, Wassily Hoeffding, Robert G. Hoffmann, John F. Hofmann, David Hogben, Paul G. Homeyer, Robert Hooke, Theodore Wright Horner, William H. Horton, Daniel G. Horvitz, Professor Harold Hotelling, Earl E. Houseman, David R. Howes, Walter W. Hoy, Cyril J. Hoyt, John David Hromi, Harry M. Hughes, J. Stuart Hunter, David V. Huntsberger, Benjamin Jackson, John L. Jaech, T. A. Jeeves, Milton Vernon Johns, Jr., Howard L. Jones, Hyman B. Kaitz, Samuel Karlin, Abraham, E. Karp, Marvin A. Kastenbaum, Leo Katz, Mort Keats, Oscar Kempthorne, Robert W. Kennard, George H. Kennedy, Bradford F. Kimball, Edgar P. King, Cal J. Kirchen, Leslie Kish, Truman L. Koehler, Martin Krakowski, Clyde Y. Kramer, William C. Krumbein, William Kruskal, Morton Kupperman, George M. Kuznets, Mrs. Helen Humes Lamale, Donald E. Lamphiear, Fred C. Leone, Howard Levene, Alfred Lieberman, Gerald J. Lieberman, Gilbert Lieberman, Jacob E. Lieberman, Julius Lieblein, Benjamin Lipstein, Stuart P. Lloyd, Frederic M. Lord, Eugene Lukacs, Bob Lundegard, George F. Lunger, John Hans MacKay, William G. Madow, Ralph A. Maggio, Clifford Joseph Maloney, Joseph Mandelson, Henry Berthold Mann, Eli S. Marks, Robert H. Matthias, Philip John McCarthy, Duncan C. McCune, Harlley Ellsworth McKean, Paul M. Meier, Margaret Merrell, W. Jay Merrill, Herbert A. Meyer, Paul D. Minton, Robert Mirsky, Sutton Monro, Alex M. Mood, Roger H. Moore, Donald Frank Morrison, Milton NMI Morrison, Norman Morse, Jack Moshman, Frederick Mosteller, Mervin E. Muller, R. B. Murphy, Jack Nadler, L. F. Nanni, Raymond Nassimbene, Joseph Anthony Navarro, August A. Carl Nelson, Jr., Peter E. Ney, S. I. Neuwirth, George E. Nicholson, Monroe L. Norden, Jack I. Northam, Horace W. Norton, Aloysius Joachim O'Connor, Junjiro Ogawa, Ingram Olkin, Paul S. Olmstead, Bernard Ostle, Donald B. Owen, William R. Pabst, Jr., Nancy S. Parker, Dr. Ellis F. Parmenter, Emanuel Parzen, John F. Pauls, Robert Nixon Pendergrass, B. E. Phillips, Eugene W. Pike, Edwin James George Pitman, Aloysius J. Polaneczky, Bruce P. Price, Ronald Pyke, Dana Edward Anthony Quade, Lila Knudsen Randolph, Herman Ravitch, Stanley Reiter, Elmer Edwin Remmenga, G. J. Resnikoff, William L. Roach, Jr., Spencer W. Roberts, Herbert Robbins, Douglas S. Robson, Robert Roeloffs, Harry M. Rosenblatt, Joan R. Rosenblatt, Murray Rosenblatt, Irving Roshwalb, S. N. Roy, Herman Rubin, David Rubinstein, Phillip Justin Rulon, Marion M. Sandomire, F. E. Satterthwaite, Sam Cundiff Saunders, Leonard S. Savage, Marvin A. Schneiderman, Seymour Max Selig, Richard H. Shaw, Sidney Shtulman, Walt R. Simmons, Rosedith Sitgreaves, John H. Smith, Thaddeus L. Smith, Jean E. Smolak, Milton Sobel, Herbert Solomon, Paul N. Somerville, Frederick

A. Sorensen, Melvin Dale Springer, John J. Stansbrey, James Hall Stapleton, Roberg G. D. Steel, Arthur Stein, Frederick F. Stephan, Theodor D. Sterling, John N. Stewart, Ray B. Stiver, Jr., David S. Stoller, Samuel A. Stouffer, Jacques St. Pierre, Hale C. Sweeny, Zen Szatrowski, Robert J. Taylor, James G. C. Templeton, Benjamin J. Tepping, Milton E. Terry, Earl A. Thomas, William Alfred Thompson, Jr., George W. Thomson, Leo J. Tick, John W. Tukey, Malcolm E. Turner, Hubertus Robert Van der Vaart, Herman W. Von-Guerard, Helen M. Walker, David L. Wallace, W. Allen Wallis, John E. Walsh, Louis Weiner, Harry Weingarten, Irving Weiss, Phillips Whidden, Alfred G. Whitney, D. Ransom Whitney, John M. Wiesen, Frank Wilcoxon, Martin B. Wilk, Samuel S. Wilks, John W. Wilkinson, Evan James Williams, Gregory Williams, Myron J. Willis, Russell Lowell Wine, Gerald Winston, Max A. Woodbury, G. Stanley Woodson, Charles Ashley Wright, Charles W. Wright, William J. Youden, Marvin Zelen, John Arthur Zoellner.

Visiting Foreign Mathematicians

The following list of visiting foreign mathematicians has been received from the Division of Mathematics, National Academy of Sciences—National Research Council. The information given is, in order, the name, home country, host institution, and period of visit; AY stands for academic year 1957–1958.

Adams, John F.—U. K.—Institute for Advanced Study—AY; Adem, Jose—Mexico—Princeton University—Feb. 1958–June 1958; Akizuki, Yasuo—Japan—University of Chicago—Oct. 1, 1957–June 30, 1958; Albertoni, Sergio—Italy—New York University—Sept. 1957–Feb. 1958; Andreotti, Aldo—Italy—Institute for Advanced Study (Sept. 30, 1957–Dec. 20, 1957), Princeton University (Feb. 1958–June 1958)—Sept. 30, 1957–June 1958; Azumaya, Goro—Japan—Yale University—Sept. 1956–Sept. 1958; Beale, E. M. L.—U. K.—Princeton University—Jan. 1958–Dec. 1958; Birch, Bryan J.—U. K.—Princeton University—Sept. 1957–June 1958; Björck, Göran—Sweden—Institute for Advanced Study—AY; Bofinger, Victor J.—Australia—North Carolina State College—June 1957–April 1958; Burgers, Johannes M.—Netherlands—American University, National Bureau of Standards—Oct. 11, 1956–Oct. 1957; Carleson, Lennart—Sweden—Massachusetts Institute of Technology—Sept. 1957–Jan. 31, 1958; Cartier, Pierre—France—Institute for Advanced Study—AY; Chakravorti, J. G.—India—Brown University—AY; Chand, Uttam—India—Boston University—Jan. 1958–May 1958; Cohen, Daniel E.—U. K.—Princeton University—Sept. 1957–June 1958; Copson, E. T.—Scotland—Stanford University—Week-Feb. 1958; Corsten, L. C. A.—Netherlands—University of North Carolina—Sept. 1957–June 1958; Dedecker, Paul—Belgium—Institute for Advanced Study—Sept. 30, 1957–Dec. 20, 1957; Delsarte, Jean—France—University of Maryland—Apr. 1957–July 1957; Deny, Jacques—France—Institute for Advanced Study—Sept. 30, 1957–Dec. 20, 1957; de Rham, Georges—Switzerland—Institute for Advanced Study—AY; Dold, Albrecht—Germany—Institute for Advanced Study—Sept. 1956–Aug. 1958; Dvoretzky, Aryeh—Israel—Institute for Advanced Study—AY; Edwards, David A.—U. K.—Yale University—Sept. 1956–Sept. 1958; Ewald, Guenther—Germany—Michigan State University—Sept. 1957–June 1958; Festa, Rudolf—Austria—University of Alabama, 1956–57 (State College of Washington 1957–58)—Sept. 1956–Sept. 1958; Festa, Erika—Austria—State College of Washington (Sept.–Dec. 1957)—Sept. 1956–Sept. 1958; Foguel, Shaul—Israel—New York University—AY; Fröhlich, A.—U. K.—University of Virginia—Feb. 1958–June 1958; Gamblen, Frank—Australia—University of Kansas (Sept. 4, 1957–Feb. 1, 1958), Educational Testing Service, Princeton, N. J. (Feb. 1, 1958–June 1958)—Sept. 4, 1957–June 1958; Gautschi, Walter—Switzerland—American University, National Bureau of Standards—Oct. 1955–Sept. 1958; Ghaffari, A.

—Iran—American University, National Bureau of Standards—Sept. 1956–Sept. 1958; Goldner, Siegfried—Union of South Africa—New York University—AY; Grauert, Hans—Germany—Institute for Advanced Study—AY; Grenander, U.—Sweden—Brown University—AY; Griffiths, Hubert B.—U. K.—Institute for Advanced Study—Sept. 1956–July 1958; Guttman, Irwin—Canada—Princeton University—Sept. 1957–June 1958; Hano, Jun-ichi—Japan—University of Washington—Sept. 15, 1957–June 15, 1958; Harrop, Ronald—U. K.—Pennsylvania State University—Aug. 1957–Aug. 1958; Hellman, Olavi B.—Finland—University of California, Los Angeles—July 1956–July 1958; Helmberg, Gilbert—Austria—University of Washington—October 1, 1957–June 15, 1958; Hervé, Michel—France—Institute for Advanced Study—Sept. 30, 1957–Dec. 20, 1957; Hirsch, Guy—Belgium—Massachusetts Institute of Technology—Feb. 1, 1958–June 15, 1958; Hitotumatu, Sin—Japan—Stanford University—Sept. 1, 1957–June 30, 1958; Hüsser, Rudolph—Switzerland—University of California, Los Angeles, AY; Izumi, Shin-ichi—Japan—University of Chicago and Northwestern University (Aug. 1, 1957–May 31, 1958), Princeton University (Oct. 1957–Dec. 1957)—Aug. 1957–May 1958; Kato, Tosio—Japan—New York University—Sept.-Oct. 1957; Kawata, T.—Japan, Princeton University—Sept. 1957–March 1958; Kitawaga, T.—Japan—Princeton University—Sept. 1957–March 1958; Klingenberg, Wilhelm—Germany—Institute for Advanced Study—AY; Laasonen, V. Pentti J.—Finland—University of California, Los Angeles—July 1956–Aug. 1958; Lacombe, Daniel L. M.—France—Institute for Advanced Study—Oct. 1957–Aug. 1958; Lehto, Olli E.—Finland—Institute for Advanced Study—AY; Leopoldt, Heinrich W.—Germany—Institute for Advanced Study—Sept. 1956–Aug. 1958; Leray, Jean—France—Institute for Advanced Study—Sept. 30, 1957–Dec. 20, 1957; Lions, Jacques—France—University of Kansas—Feb. 1957–Aug. 1958; Longuet-Higgins, Michael S.—U. K.—Massachusetts Institute of Technology—Feb. 1, 1958–June 15, 1958; Lorenzen, Paul P. W.—Germany—Institute for Advanced Study—Sept. 1957–June 1958; Lucas, John R.—U. K.—Princeton University—Sept. 1957–June 1958; Lumer, Günter—Uruguay, University of Chicago—Oct. 1, 1957–Sept. 30, 1958; Mallows, Colin L.—U. K.—Princeton University—Sept. 1957–Sept. 1958; Mardešić, Sibe—Jugoslavia—Institute for Advanced Study—AY; Martin, Alfred I. —U. K.—Institute for Advanced Study—AY; Masani, Pesi—India—Massachusetts Institute of Technology—Harvard University—Sept. 16, 1957–June 15, 1958; Message, Philip J.—U. K.—Yale University—Sept. 1957–Sept. 1958; Milne-Thomson, L. M.—U. K.—Brown University—Sept. 1956–June 1958; Mixner, Joseph—Germany—New York University—Aug.-Oct. 1957; Møller, Christian—Denmark—Carnegie Institute of Technology—Sept. 1957–Feb. 1958; Nachbin, Leopoldo—Brazil—University of Chicago—Oct. 1, 1956–July 31, 1958; Nagata, Masayoshi—Japan—Harvard University—Sept. 1957–Sept. 1958; Nieminen, Toivo E.—Finland—New York University—Aug. 1957–June 1958; Ogawa, Junjiro—Japan—Institute of Statistics, University of North Carolina—Sept. 1956–Aug. 31, 1958; Olver, F. W. J.—U. K.—National Bureau of Standards—Sept. 30, 1957–Sept. 1958; O'Meara, Onorato T.—South Africa—Institute for Advanced Study—AY; Ono, Katsuzi—Japan—Massachusetts Institute of Technology—Sept. 1957–June 1958; Ostrowski, Alexander M.—Switzerland—American University, National Bureau of Standards—Oct. 10–31, 1957; Papakyriakopoulos, C. D.—Greece—Institute for Advanced Study—June 3, 1955–June 1958; Peixoto, Mauricio M.—Brazil—Princeton University—Sept. 1957–June 1958; Pfluger, Albert—Switzerland—Stanford University—Oct. 1, 1957–Mar. 30, 1958; Pucci, Carlo—Italy—University of Maryland—Sept. 1, 1956–July 1958; Puppe, Dieter—Germany—Institute for Advanced Study—Sept. 1957–June 1958; Rieger, Georg J.—Germany—University of Maryland—Sept. 1956–Aug. 1957; Riesz, Marcel—Sweden—University of Maryland—Oct. 1, 1957–Dec. 31, 1957; Robinson, Leslie R. B.—U. K.—Harvard University—Sept. 1957–Sept. 1953; Rogosinski, Werner W.—U. K.—University of Colorado—Sept. 1957–Sept. 1958; Rohrbach, Hans—Germany—University of North Carolina—Sept. 1957–June 1958; Room, T. G.—Australia—Institute for Advanced Study (AY); Princeton University (Feb. 1958–June 1958)—Sept. 1957–June 1958; Roseau, Maurice—France—New York University

—Sept. 1957-Sept. 1958; Sawyer, W. W.—U. K.—University of Illinois—Feb. 1957-Indefinite; Schopf, Andreas—Switzerland—American University-National Bureau of Standards—Sept. 30, 1957-Oct. 1958; Scriba, J. Cristoph—Germany—University of Kentucky—Sept. 1957-June 1958; Selberg, Sigmund—Norway—University of Colorado—Sept. 1957-May 1958; Serre, Jean-Pierre—France—Institute for Advanced Study—Sept. 30, 1957-Dec. 20, 1957; Skolem, Thoralf A.—Norway—University of Notre Dame—Sept. 1957-June 1958; Stoll, Wilhelm—Germany—Institute for Advanced Study—AY; Tamagawa, Tsuneo—Japan—Institute for Advanced Study (Sept. 1955-Jan. 1957 and Jan. 13, 1958-Apr. 11, 1958), Johns Hopkins University (Jan. 1957-Jan. 1958)—Sept. 1955-Apr. 1958; Tomonaga, Yasuro—Japan—University of Washington—Oct. 15, 1957-June 15, 1958; Valpola, Veli Kustaa—Finland—University of California (2 months), Princeton University (4 months)—Nov. 1957-April 1958; van der Vaart, H. R.—Netherlands—North Carolina State College—Jan. 1957-Jan. 1958; Villamayor, Orlando—Argentina—Institute for Advanced Study—Jan. 1, 1957-Dec. 31, 1957; Waelbroeck, L.—Belgium—Institute for Advanced Study—AY; Watson, Geoffrey S.—Australia—North Carolina State College (4 months), Princeton University (5 months)—April 1958-Dec. 1958; Williams, Robert M.—U. K.—Princeton University—Sept. 1957-June 1958; Wolff, Emil—U. K.—New York University—June-Nov. 1957; Yamamuro, Sadayuki—Japan—Institute for Advanced Study—Sept. 1956-Sept. 1958; Yevdjevich, V. M.—Yugoslavia—American University, National Bureau of Standards—AY; Yüksel, H.—Turkey—Brown University—AY; Zadunaisky, Pedro—Argentina—Watson Scientific Computing Laboratory—Feb. 1, 1957-Jan. 31, 1958; Agudo, F. R. D.—Portugal—University of California, Berkeley Fall 1957; Baayen, Pieter C.—Netherlands—University of California, Berkeley—AY; Fary, Istvan—Canada (Hungary)—University of California, Berkeley Jan.-June 1958; Festa, Erika—Austria—State College of Washington Sept.-Dec. 1957; Lightstone, A. H.—Canada—University of California, Berkeley—AY; Littlewood, J. E.—England—University of California, Berkeley Sept. 27-Dec. 18, 1957; Poulsen, Ebbe T.—Denmark—University of California, Berkeley July 1957-June 1958; Specker, Ernst Paul—Switzerland—Cornell University Feb. 10-Sept. 1958; Szmielew, Wanda—Poland—University of California, Berkeley—AY.

Committee on Mathematical Tables

Since its organization early in 1956, the Institute of Mathematical Statistics' Committee on Mathematical Tables has been concerned with the problems associated, either directly or indirectly, with the computation of mathematical tables of interest to statisticians. The committee's function is threefold:

(i) To gather information relating to the tabulation of functions of interest to statisticians.

(ii) To advise on the need for, and preparation of, statistical tables.

(iii) To determine the availability of and coordinate the distribution of free time on high speed digital computers for the computation of statistical tables.

In order to fulfill its function, the committee investigated the interests and needs of the Institute membership concerning statistical tables and as a result set up nine subcommittees covering the areas of greatest interest. To date the activities of these subcommittees have been directed primarily towards the preparation of bibliographies in their individual fields. The subcommittees, along with their chairmen, are listed below.

1. Chi-Square,
   W. Kruskal, University of Chicago
2. t-Distribution (Univariate and Multivariate),
   C. W. Dunnett, American Cyanamid Company, Pearl River, New York.
3. Studentized Range,
   A. H. Bowker, Stanford University
4. F-Distribution (Incomplete-Beta, Binomial),
   E. E. Cureton, University of Tennessee.
5. Hypergeometric Distribution (*Not* the hypergeometric function),
   W. Kruskal, University of Chicago.
6. Polyvariate Normal Distribution, including latent roots,
   G. P. Steck, Sandia Corporation, Albuquerque, New Mexico
7. Availability of Simple Techniques,
   Chairman not appointed.
8. Annals Supplement (Of statistical tables),
   J. W. Tukey, Bell Telephone Laboratories, Murray Hill, New Jersey
9. Computing Facilities and Cost-Free Machine Time,
   F. C. Leone, Case Inst. of Tech.

Additional information on any of the above activities may be obtained from the chairman of the Committee on Mathematical Tables, D. B. Owen, Sandia Corporation, Albuquerque, New Mexico, or from any of the subcommittee chairmen. Anyone having time on a digital computer which may be made available on a cost-free basis to persons desiring to compute tables of general interest is invited to contact the chairman of subcommittee 9.

I. *List of Members of the IMS Committee on Mathematical Tables*

Chairman

Dr. D. B. Owen, Division 5125, Sandia Corporation, Albuquerque, New Mexico

Secretary and Vice Chairman

Dr. G. P. Steck, Division 5125, Sandia Corporation, Albuquerque, New Mexico

Members

Professor R. L. Anderson, Institute of Statistics, North Carolina State College, Raleigh, North Carolina

Professor A. H. Bowker, Department of Statistics, Stanford University, Stanford, California

Professor E. E. Cureton, 1846 Prospect Pl., S.E., Knoxville 15, Tennessee

Professor W. J. Dixon, University of California, Department of Preventive Medicine and Public Health, Medical Center, Los Angeles 24, California

Mr. C. W. Dunnett, 19 Edsall Place, Nanuet, New York

Dr. Churchill Eisenhart, Chief, Statistical Engineering Laboratory, National Bureau of Standards, Washington 25, D. C.

Dr. J. A. Greenwood, 16 Garfield Street, Cambridge 38, Massachusetts

Professor H. O. Hartley, Statistical Laboratory, Iowa State College, Ames, Iowa

Professor William Kruskal, Committee on Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois

Professor Fred C. Leone, Director, Statistical Laboratory, Case Institute of Technology, 10900 Euclid Avenue, Cleveland 24, Ohio
Professor Dan Teichroew, Graduate School of Business, Stanford University, Stanford, California
Dr. John W. Tukey, Bell Telephone Laboratories, Murray Hill, New Jersey
Professor M. A. Woodbury, 401 W. 205th Street, New York 34, New York
Dr. Marvin Zelen, Statistical Engineering Laboratory, National Bureau of Standards, Washington 25, D. C.

Ex officio members

Professor L. J. Savage, Eckhart Hall, University of Chicago, Chicago 37, Illinois.
Professor Jacob Wolfowitz, Dept. of Math, Cornell University, Ithaca, New York

II. *Subcommittees of the IMS Committee on Mathematical Tables*

1. *Chi-square*
   *William Kruskal
   *Dan Teichroew

2. *t-distributions* (univariate and multivariate)
   *C. W. Dunnett, Chairman
   Dr. H. A. David, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia
   Dr. S. S. Gupta, Department of Mathematics, University of Alberta Edmonton, Alberta, Canada
   *H. O. Hartley
   Professor E. S. Keeping, Math Department, University of Alberta, Edmonton, Alberta, Canada
   Professor C. F. Kossack, Math Department, Purdue University, West Lafayette, Indiana
   Dr. A. M. Mood, General Analysis Corporation, 11753 Wilshire Boulevard, West Los Angeles, California
   J. B. Rabin, Sen. Computer Analyst, Burroughs Corporation, 1505 Sycamore Avenue, Willow Grove, Pennsylvania
   Dr. M. Sobel, Bell Telephone Lab., 555 Union Boulevard, Allentown, Pa.
   *Dan Teichroew

3. *Studentized range*
   *A. H. Bowker, Chairman,
   Cuthbert Daniel, 116 Pinehurst Avenue, New York 33, New York
   Prof. W. T. Federer, Cornell University, Ithaca, New York
   *H. O. Hartley
   Professor G. E. Noether, Math Department, Boston University, 725 Commonwealth Ave., Boston 15, Massachusetts

4. *F-distribution* (incomplete-beta, binomial)
   *E. E. Cureton, Chairman,
   *R. L. Anderson
   P. C. Cox, 1904 Idaho Avenue, Las Cruces, New Mexico
   Professor David Durand, 50 Memorial Drive, Cambridge 39, Massachusetts
   *J. A. Greenwood
   *H. O. Hartley
   Gunnar Kulldorff, University of Lund, Lund, Sweden, Malmgatan 16, Malmo, Sweden
   *Dan Teichroew

---

* Member of the parent committee. See List I for address.

**H. F. Trotter

Prof. J. G. Wendel, Math Department, University of Michigan, Ann Arbor, Michigan

5. *Hypergeometric distribution* (*not* the hypergeometric function)
   *William Kruskal, Chairman,
   Professor Leo Katz, Department of Statistics, Michigan State University, East Lansing, Michigan
   Dr. E. F. Kimball, N. Y. State Pub. Service Commission; 20 Mayfair Drive, Slingerlands, New York
   Professor G. J. Lieberman, Department of Statistics, Stanford University, Stanford, California
   Roger H. Moore, Los Alamos Scientific Laboratory, 3448A Orange, Los Alamos, New Mexico
   J. M. Wiesen, 1308 Arizona NE, Albuquerque, New Mexico

6. *Polyvariate normal*, including latent roots
   *G. P. Steck, Chairman
   Professor T. W. Anderson, Center for Advanced Study in the Behavioral Sciences, 202 Junipero Serra Blvd., Stanford, California
   P. C. Cox, (See Subcommittee 4 for address)
   *C. W. Dunnett
   S. S. Gupta, (See Subcommittee 2 for address)
   Professor Ingram Olkin, Department of Statistics, Michigan State University, East Lansing, Michigan
   *D. B. Owen
   M. Sobel, (See Subcommittee 2 for address)
   *Max A Woodbury

7. *Availability of simple techniques*
   Chairman position open.
   Professor R. A. Bradley, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia
   *E. E. Cureton
   *W. J. Dixon
   *Churchill Eisenhart
   Dr. T. A. Lamke, Bu. of Res., Iowa State Teachers College, Cedar Falls, Iowa
   Professor S. B. Littauer, Columbia University, New York 27, New York
   **H. R. Watkins

8. *Annals Supplement*
   *J. W. Tukey, Chairman
   J. S. Barnes, John Wiley & Sons Inc., 440 Fourth Avenue, New York 16, N. Y.
   *A. H. Bowker
   *Churchill Eisenhart
   Dr. T. E. Harris, The RAND Corporation, 1700 Main Street, Santa Monica, Calif.
   *D. B. Owen
   *Dan Teichroew

9. *Cost-Free Machine Time*
   *F. C. Leone, Chairman
   Professor J. W. Hamblen, Computing Center, Oklahoma State University, Stillwater, Oklahoma
   W. H. Horton, Materials Engineering Department, Westinghouse Electric Corp., East Pittsburgh, Pa.

** Address unknown.

G. F. Lunger, Program Planning Department, Remington Rand Univac, St. Paul 16, Minnesota

Dr. H. A. Meyer, Director Statistical Laboratory, University of Florida, Gainesville, Florida

*Max A. Woodbury

REPORT OF THE LOS ANGELES MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The Western Region Meeting, seventy-fifth meeting of the Institute of Mathematical Statistics, was held at the Los Angeles Campus of the University of California on December 27–28, 1957. Two sessions were joint with the American Statistical Association and the Institute for Management Sciences. A Special Invited Address was given by Claude Shannon, "Some Asymptotic Estimates for Sums of Random Variables." The following 69 members of the Institute attended:

I. J. Abrams, T. W. Anderson, H. L. Ang, Leo A. Aroian, C. B. Bell, D. L. Bentley, Allan Birnbaum, Colin R. Blyth, Charles Boll, Julien L. Borden, A. H. Bowker, J. V. Breakwell, Bernice Brown, Herman Chernoff, Edward P. Coleman, L. M. Court, Edwin L. Crow, W. J. Dixon, Olive Jean Dunn, H. P. Edmundson, Bob E. Ellison, T. S. Ferguson, Evelyn Fix, Martin Fox, A. V. Gafarian, Edward Gammon, Norman R. Garner, E. J. Gilbert, F. A. Graybill, Wm C. Guenther, D. Guthrie, Jr., T. E. Harris, P. G. Hoel, John F. Hofmann, John M. Howell, Arnljot Høyland, Patricia Inman, M. V. Johns, Jr., R. F. Link, Albert Madansky, Craig A. Magwire, F. Massey, M. R. Mickey, O. B. Moan, Roger A. Moore, Paul B. Moranda, James Pachares, Emanuel Parzen, M. P. Peisakoff, H. H. Peterson, Ron Pyke, Roy Radner, F. C. Reed, David Rothman, Marion M. Sandomire, Henry Scheffé, E. M. Scheuer, Franklin Sheehan, Bernard Sherman, M. M. Siddiqui, Paul N. Somerville, D. Stoller, Fred H. Tingey, Howard G. Tucker, John W. Tukey, H. W. von Guérard, John E. Walsh, Louis H. Wegner, Bryan Wilkinson.

The program of the meeting was as follows:

Friday, December 27, 1957

8:45–12:00 a.m. Statistics in the Management Sciences
(Joint Session with The Institute of Management Sciences)

Chairman: M. R. Mickey, Jr., The RAND Corporation.

1. *The Portfolio Selection Problem*, Harry Markowitz, The RAND Corporation.
2. *On the Stochastic Theory of Inventory*, I. J. Abrams, The Ramo-Wooldridge Corporation.
3. *Inventory Control Problems of Shipboard Supplies*, Mina H. Gourary, George Washington University. (Read by Bernice B. Brown, The RAND Corporation).
4. *Demand for and Allocation of Engineering Personnel*, Rajendra Kashyap (introduced by H. W. von Guerard) and Hermann W. von Guerard, Lockheed Aircraft Corporation.

1:45–2:15 p.m. Special Invited Paper

Chairman: Thomas S. Ferguson, University of California, Los Angeles
Some Asymptotic Estimates for Sums of Random Variables
Claude Shannon, MIT, The Center for Advanced Study in the Behavorial Sciences, and Bell Telephone Laboratories.

2:30–5:00 p.m. Industrial Applications of Statistics

(Joint Session with the American Statistical Association)

Chairman: John F. Hofmann, Systems Laboratories Corporation

1. *Some Applications of Experimental Design in Industry*, Alex M. Mood and Paul Somerville, General Analysis Corporation.
2. *The Fitting of a Polynomial Form to a Function of Several Variables by the Use of Orthogonal Latin Squares*, N. M. Peterson, Convair, Fort Worth.
3. *Confidence Intervals for the Reliability of Multi-Stage Systems*, William C. Hoffman, The RAND Corporation.
4. *A Model for Depicting Fatigue*, Irvin Whiteman, General Analysis Corporation (introduced by A. M. Mood).
5. *Long Range Planning for Manufacturing*, Glen Ghormley, Cannon Electric Company.

Saturday, December 28, 1957

8:45–10:45 a.m. Stochastic Processes Applied to Medicine and Public Health

Chairman: Frank J. Massey, University of California, Los Angeles

1. *Replication Versus Increasing Observation Points in the Estimation of Regression for Growth Type Data*, Paul G. Hoel, University of California, Los Angeles.
2. *The Identifiability Problem for Functions of Finite Markov Chains*, Edgar John Gilbert, University of California, Berkeley.

11:00–12:00 a.m. Invited Address

Chairman: Roger A. Moore, The Ramo-Wooldridge Corporation

1. *Statistical Theory of Some Quantal Response Models*, Allen Birnbaum, Columbia University.

1:45–2:45 p.m. Invited Address

Chairman: O. B. Moan, Lockheed Aircraft Corporation

1. *Experiments with Mixtures*, Henry Scheffé, University of California, Berkeley.

3:00–5:00 p.m. Contributed Papers

Chairman: Richard F. Link, Oregon State College

1. *Non-parametric Multiple-decision Procedures for Selecting That One of K Populations which has the Highest Probability of Yielding the Largest Obser-*

vation. (Preliminary Report). Robert Bechhofer, Cornell University, and Milton Sobel, Bell Telephone Laboratories. (By Title)

2. *The Asymptotic Efficiency of Friedman's Chi-square$_r$-test* (χ_r^2-test). Ph. van Elteren, Mathematical Centre, Amsterdam. (By title)
3. *Least-squares Estimation when Residuals are Correlated.* M. M. Siddiqui, University of North Carolina.
4. *A Property of Additively Closed Families of Distributions.* Edwin L. Crow, Boulder Laboratories, National Bureau of Standards.
5. *Determining Sample Size for a Specified Width Confidence Interval.* Franklin A. Graybill, Oklahoma State University.
6. *Nonparametric Estimation of Sample Percentage Point Standard Deviation.* John E. Walsh, Lockheed Aircraft Corporation.
7. *On the Structure of Distribution-free Statistics.* C. B. Bell, Xavier University of Louisiana and Stanford University.
8. *Estimation of the Location of a Discontinuity in Density.* J. V. Breakwell, Lockheed Missile Systems Division, Palo Alto, and H. Chernoff, Stanford University.
9. *On the Supremum of the Poisson Process.* Ronald Pyke, Stanford University.

EVELYN FIX
Associate Secretary

REPORT OF THE EDITOR OF THE ANNALS FOR 1957

During the year ending August 1, 1957, more new manuscripts were received by the *Annals*, totaling more manuscript pages, than in any previous year. A consequent increased requirement for printing is anticipated for the coming year, and the Council has authorized a 1958 volume of 1300 pages.

The 1957 volume, totaling 1098 pages, contained 105 papers and notes. The increased size authorized by the Council in the past two years made it possible to keep the backlog at less than half an issue during 1957.

The *Annals* is indebted to its staff of Cooperating Members, who do much of the refereeing, and to the following people who have generously given refereeing assistance: T. W. Anderson, P. Armitage, E. W. Barankin, M. S. Bartlett, R. Blumenthal, R. Bechhofer, G. E. P. Box, L. Breiman, D. L. Burkholder, S. Chandrasekhar, W. G. Cochran, W. S. Connor, L. Cote, S. L. Crump, F. N. David, M. D. Donsker, R. Dorfman, A. Duncan, M. Dwass, B. Epstein, P. Erdös, T. S. Ferguson, L. J. Folks, E. J. Gilbert, I. J. Good, L. Goodman, F. Graybill, U. Grenander, S. S. Gupta, J. F. Hannan, M. H. Hansen, W. Hoeffding, P. G. Hoel, H. Hotelling, A. T. James, N. L. Johnson, E. S. Keeping, J. H. B. Kemperman, D. G. Kendall, M. G. Kendall, H. Kesten, E. L. Lehmann, J. Leiblein, R. Leipnik, M. Loève, E. Lukacs, J. McGregor, A. Madansky, W. G. Madow, M. R. Mickey, A. M. Mood, P. A. P. Moran, F. Mosteller, J. Moyal, R. W. Murphy, M. Newman, I. Olkin, E. Parzen, R. L. Plackett, J. Pratt,

R. Pyke, C. R. Rao, D. Ray, G. E. H. Reuter, J. Riordan, M. Rosenblatt, H. L. Royden, H. Rubin, J. Sacks, I. R. Savage, H. Scheffé, E. L. Scott, J. F. Scott, R. Sitgreaves, C. Streibel, R. F. Tate, D. Teichroew, A. J. Thomasian, H. Trotter, J. W. Tukey, D. L. Wallace, J. E. Walsh, B. L. Welch, L. Wegner, R. A. Wijsman, D. M. G. Wishart, G. Zyskind.

Many thanks are due Ann Greene, Dorothy Stewart, and Margaret Wray, for handling the taxing work of the editorial office.

T. E. Harris
Editor

December 26, 1957

SUMMER SESSIONS AT BERKELEY, CALIFORNIA

The 1958 summer program in the Department of Statistics of the University of California, Berkeley, California, will consist of two sessions: June 16 to July 26 and July 28 to September 6. The faculty of the summer sessions will include Professor U. S. Nair of Travancore University in India, Dr. F. N. David of University College in London, and Professors David Blackwell, Evelyn Fix, Joseph L. Hodges, Jr., and J. Neyman of the Department of Statistics of the University of California, Berkeley. The program will include two undergraduate courses in each session, and two research seminars, one in statistical problems of health and one in the statistical study of structural relations in the physical sciences.

PUBLICATIONS RECEIVED

Chow, G. C., *Demand for Automobiles in the United States*, North-Holland Publishing Company, Amsterdam, (1957) v-110 pp., $10.00.

Economica, *Revista de la Facultad de Ciencias Economicas*, Publicacion Trimestral, Diagonal 77-4 Y 5, La Plata, Buenos Aires, Argentina.

Contributions to the Theory of Games, Volume III, Edited by M. Dresher, A. W. Tucker, P. Wolfe, Princeton, New Jersey, Princeton University Press, 1957 Annals of Mathematics Studies Number 39.

ECONOMETRICA

Journal of the Econometric Society

Contents of Vol. 26, No. 1 - January, 1958

N. GEORGESCU-ROEGEN .... Threshold in Choice and the Theory of Demand
F. HAHN .... Gross Substitutes and the Dynamic Stability of General Equilibrium
K. J. ARROW .... Utilities, Attitudes, Choices: A Review Note
A. S. MANNE .... A Linear Programming Model of the U.S. Petroleum Refining Industry
J. TOBIN .... Estimation of Relationships for Limited Dependent Variables
M. BECKMANN .... Some Aspects of the Airline Reservations Problem
P. CARRE ...... Tentative de détermination empirique de fonctions de production pour les pays industriels
H. WAGNER .... A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equation
G. STUVEL ...... The Impact of Changes in the Terms of Trade on Western Europe's Balance of Payments
M. FISHER .... A Sector Model: The Poultry Industry of the U.S.A.

Economic Efficiency in Plant Operations with Special Reference to the Marketing of California Pears (French, Sammet, and Bressler). Review by F. V. Waugh.
Business Forecasting, 1956 (Abramson, Mack, and others). Review by Charles F. Roos.
Einfuhrung in die mathematische Statistik (L. Schmetterer). Review by William Feller.
International Comparisons of Real Wages (International Labour Office). Review by Kurt W. Rothschild.
Income and Wealth. Series V. (S. Kuznets, ed.) Review by J. B. D. Derksen.
Trends in Employment in the Service Industries (G. J. Stigler). Review by Colin Clark.
Zinstheorie (F. A. Lutz). Review by Joseph Aschheim.
Foundations of Productivity Analysis (Bela Gold). Review by C. F. Carter.
Rapport sur les comptes de la nation. Review by Walter Froehlich.
Automata Studies (C. E. Shannon and J. McCarthy, eds.). Review by Harry H. Goode.
Economic Progress (Léon H. Dupriez, ed.). Review by Leif Johansen.
On Economic Theory and Socialism, Collected Papers (Maurice Dobb). Review by Kenneth O. May.
Statistics: A New Approach (W. A. Wallis and H. V. Roberts). Review by Maurice Quenouille.
Die Ermittlung der wirtschaftlichen Nutzungsdauer von Anlagegüetern (Hansrudolf von Briel.) Review by Eric Schiff.
Structural Interdependences of the Economy: Proceedings of an International Conference on Input-Output Analysis (T. Barna, ed.). Review by Robert Solow.
International Economic Papers No. 4: Translations Prepared for the International Economic Association (Peacock, Turvey, Stolper, and Henderson, eds.). Review by Paul M. Sweezy.
Marketing Efficiency in Puerto Rico (Galbraith, Holton, *et al.*). Review by Lester G. Telser.
Méthodologie économique (Gilles-Gaston Granger). Review by Sten Thore.
Studies in the Economics of Transportation (Beckmann, McGuire Winsten, and Koopmans). R. M. Thrall.
Capital and Its Structure (L. M. Lachmann). Review by W. P. Yohe.

JOURNAL OF THE

ROYAL STATISTICAL SOCIETY

Series B (Methodological)

30s per part — Vol. 19 No. 2 1957 — Ann. Sub. Incl. post £3 1s 0d

CONTENTS

Problems in the Probability Theory of Storage Systems .... J. GANI.
Some Problems in the Theory of Dams .... D. G. KENDALL. (With Discussion)
Routine Analysis of Replicated Experiments on An Electric Computer.
F. YATES, M. J. R. HEALY and S. LIPTON. (With Discussion)
Minimax Procedure for choosing between Two Populations using Sequential Sampling ...... R. J. MAURICE
Sufficient Statistics, Similar Regions and Distribution-free Tests .... G. S. WATSON.
On the Use of the Normal Approximation in the Treatment of Stochastic Processes .... P. WHITTLE.
The Variance of the Mean of a Stationary Process .... E. J. HANNON.
The Comparison of Regression Variables .... E. J. WILLIAMS.
Methods of Construction and Analysis of Serially Balanced Sequences .... M. R. SAMPFORD.
The Modified Latin Square .... B. ROJAS and R. F. WHITE.
Analysis of Variance as an Alternative to Factor Analysis .... M. A. CREASY.
Some Further Results in the Non-Equilibrium Theory of a Simple Queue .... N. T. J. BAILEY.
Normal Approximation to Machine Interference with Many Repair Men .... P. NAOR.
Stationary Distributions of the Negative Experimental Type for the Infinite Dam.
J. GANI and N. U. PRABHU.

The Royal Statistical Society, 21, Bentinck Street, London, W. 1

SANKHYĀ

The Indian Journal of Statistics
Edited by P. C. Mahalanobis

Vol. 18, Parts 3 & 4, 1957

The Underworld of Probability .... Sir Ronald Fisher
On Unbiased Estimates of Uniformly Minimum Variance .... R. R. Bahadur
National Sample Survey Number Nine: Survey of Displaced Persons in the Urban Areas of the Bombay State
The Multiplex Method for Linear Programming .... Ragnar Frisch
Soldiers With Unusually Low Blood Pressure .... N. T. Mathew
Unbiased Minimum Variance Estimation in a Class of Discrete Distribution .... Jogabrata Roy and Sujit Kumar Mitra
Ordered and Unordered Estimators in Sampling Without Replacement .... M. N. Murthy
A Lower Bound to the Probability of Student's Ratio .... Saibal Kumar Banerjee

Annual Subscription: 30 rupees ($10.00), 10 rupees ($3.50) per issue.
Back Numbers: 45 rupees ($15.00) per volume; 12/8 rupees ($4.50) per issue.
Subscriptions and orders for back numbers should be sent to
STATISTICAL PUBLISHING SOCIETY
204/1 Barrackpore Trunk Road **Calcutta 35, India**

SKANDINAVISK AKTUARIETIDSKRIFT

1956 V. 39 Parts 3 – 4

Contents

INNEHALL

U. Grenander .... On the Theory of Mortality Measurement. Part II
N. F. Gjeddebaek .... Contribution to the Study of Grouped Observations. II. Loss of Information Caused by Grouping of Normally Distributed Observations
C. G. Esseen .... A Moment Inequality with an Application to the Central Limit Theorem
John Gurland .... An Inequality Satisfied by the Gamma Function
K. G Hagstroem .... Can Life Assurance—and Especially Pension Schemes—be Built on Variable Premiums? And Can Inflation Risk be Covered?
Petrus Mattsson .... New Bases for Non-cancellable Sickness Insurance in Sweden
D. E. Barton .... Neyman's ψ_k^2 Test of Goodness of Fit when the Null Hypothesis is Composite
Litteraturanmälningar
Eksamen i Forsikringsvidenskab og Statistik ved Københavns Universitet
Översikt av utländska aktuarietidskrifter
XVth International Congress of Actuaries, New York, October, 1957

Annual subscription: $5.00 per year
Inquiries and orders may be addressed to the Editor,
GRANHÄLLSVÄGEN 35, STOCKSUND, SWEDEN